TEMPORAL GENETIC DYNAMICS AND MATING TYPE INHERITANCE IN AN EXPERIMENTAL BIPARENTAL FIELD POPULATION OF *PHYTOPHTHORA CAPSICI*

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Maryn O. Carlson

August 2016

ABSTRACT

The common co-occurrence of both A1 and A2 mating types of *Phytophthora capsici* results in production of oospores, and consequently persistent, overwintering populations in many locations. To understand the dynamics of these often isolated, sexual populations, a restricted access research farm in Geneva, NY, with no prior history of *P. capsici,* was inoculated in 2008 with two isolates of opposite mating type (MT). Approximately 50 isolates were sampled each year, from 2009-13, from susceptible plant species. To provide a controlled reference for the field study, F_1 single-oospore progeny were collected from an *in vitro* cross between the same founding parents. Isolates were analyzed using genotyping-by-sequencing (GBS), which simultaneously identifies and scores single nucleotide polymorphism (SNP) markers distributed throughout the genome. Applying a genetic similarity threshold based on pairwise comparisons between replicates of the parental isolates, we identified 159 unique genotypes among the initial 232 field isolates. Declines in individual and population heterozygosity, revealed by analysis of these 159 isolates, show that over time the population underwent a generational shift; transitioning from putative F_1 in 2009-10, to mixed generational in 2011, and ultimately all inbred in 2012-13. Capitalizing on the segregation of mating type in this population, we performed a genome-wide test of allele frequency differences between isolates of opposite mating types. Charting allele frequency trajectories of SNPs in the mating type associated region, demonstrated that heterozygosity for one of the A2 founding haplotypes was consistently associated with the A2 mating type. Understanding the diverse processes influencing genetic changes in this population provide insight into the evolutionary dynamics of *P. capsici*, and may lead to improved management strategies.

BIOGRAPHICAL SKETCH

Nearing the culmination of Maryn's high school studies in Sudbury, MA, she found employment on a 50-acre vegetable farm. In this year, 2009, an unusually cool, wet summer coupled with a large source of *Phytophthora infestans* inoculum from box stores importing infected plants from the south and presence of susceptible plant cultivars, provided the ideal conditions for a tomato late blight epidemic. This serendipitous, hands-on experience with the capacity of a plant pathogen to inflict devastating crop losses, ultimately motivated Maryn's pursuit of a graduate degree in the field of Plant Pathology.

First, emerging from a high school experience rich in the humanities, Maryn naturally gravitated toward studying literature at Columbia University. After completing her freshman year, and an additional semester farming, in January of 2011, Maryn's passion for agriculture motivated her to transfer to Cornell University. While at Cornell, Maryn designed her own major, incorporating agriculture, plant science, English, and art, and continued to work on vegetable farms. In Maryn's last undergraduate semester, an introductory plant pathology course taught by *P. infestans* expert, Dr. Bill Fry, catalyzed her decision to pursue graduate study. Maryn began her graduate studies in the Fall of 2013, and performed her thesis research with Dr. Christine Smart on the oomycete vegetable pathogen *Phytophthora capsici*.

To my mentors past and present

who have inspired and challenged me in diverse pursuits

and

To my parents

who have supported me in all endeavors

ACKNOWLEDGMENTS

First, I thank my major advisor, Dr. Christine Smart for her steadfast support over the three years of my tenure as a graduate student. As a graduate student with Dr. Smart, I have become a better scientist and had the freedom to hone my research interests. I especially thank Dr. Michael Gore for his immense support and encouragement, collaboration, and for challenging me to continuously refine my research. I thank Dr. Elodie Gazave for perpetually inspiring me to think creatively, regardless of the task, for myriad population genetic conversations, and immense editorial support. In addition, I thank my committee members, Drs. Alan Collmer, Michael Gore, Niklaus Grünwald, and Michael Mazourek for their input and encouragement. I thank Dr. Bill Fry for encouraging me to pursue a graduate degree, for his continued support, and the many conversations that have transpired as a result of my frequent passage between Bradfield Hall and the Plant Science Building. I thank Amara Dunn for graciously introducing me to her research on *P. capsici* and for numerous *Phytophthora* conversations since. I thank Holly Lange for facilitating all lab and field work. I thank the undergraduate and high school students who have contributed laboratory and field work. I thank members of the Smart lab for their support and help in the lab and field. Finally, I thank members of the Gore Lab for numerous helpful conversations and useful feedback on research and presentations.

TABLE OF CONTENTS

Biographical Sketch..... iii

Dedication.....iv

Acknowledgements.....v

Table of Contents.....vi

List of Figures.....viii

List of Tables.....ix

Chapter 1: A (Brief) Historical Perspective: Oomycetes are not fungi.....1

References.....5

Chapter 2: Temporal Genetic Dynamics in an Experimental Biparental Field Population of *Phytophthora capsici*.....7

Introduction.....7

Results.....10

Discussion.....31

Materials and Methods.....36

References.....45

Conclusion.....51

Appendix A: Supplemental Text.....56

A.1 Correcting for mitotic LOH in ME estimates.....56

A.2 Evidence for LOH events.....56

A.3 Mating type associated SNPs in the field inbred subpopulation.....58

Appendix B: Supplemental Tables.....59

Appendix C: Supplemental Figures.....67

LIST OF FIGURES

Figure 2.1: Distribution of the mating type of each isolate by year…….....13

Figure 2.2: Population structure in the biparental field population…….....15

Figure 2.3 Generational shift in the field population.....17

Figure 2.4 Population heterozygosity and allele frequency distributions.....18

Figure 2.5. Mendelian errors distinguish F_1 and inbred isolates in the field.....22

Figure 2.6. Regions of differentiation between the field F_1 and inbred isolates.....24

Figure 2.7. Allele frequency differences between isolates of opposite mating types.....27

Figure 2.8. Segregation of SNPs in the mating type region.....29

SUPPLEMENTARY FIGURES

Figure C.S1. SNP and individual filtering pipeline.....67

Figure C.S2. Relationship between sequencing coverage and heterozygosity.....68

Figure C.S3. Skewed allele depth ratios in two isolates suggest ploidy variation.....69

Figure C.S4. Comparing pruned and unpruned data sets.....70

Figure C.S5. Minor allele frequency distributions for the field and *in vitro* populations.....71

Figure C.S6. Relationship between the number of SNPs in each scaffold and the incidence of Mendelian error enriched SNPs.....72

Figure C.S7. Differentiation between the *in vitro* F_1, field F_1 and field inbred subpopulations.....73

Figure C.S8. Regions of differentiation between the *in vitro* F_1 and field F_1 were associated with loss of heterozygosity events in the parental cultures.....74
Figure C.S9. Principal component analysis in scaffolds pertaining to regions of interest.....75
Figure C.S10. Phase diagrams for R-26.....76
Figure C.S11. Phase diagrams for ROI-1.....78
Figure C.S12. Phase diagrams for R-35.....80
Figure C.S13. Phasing the replicates of the parental isolates.....82
Figure C.S14. Linkage disequilibrium in regions of interest in the field inbred subpopulation.....83
Figure C.S15. Allele frequency differences between isolates of opposite mating types in the *in vitro* F_1, field F_1, and field inbred subpopulations.....84
Figure C.S16. Principal component analysis of mating type associated SNPs.....85
Figure C.S17. Heterozygosity in the mating type region.....86
Figure C.S18. Heterozygosity excess in the mating type region.....87

LIST OF TABLES

Table 2.1. Selfing *in vitro* and in the field.....20

SUPPLEMENTARY TABLES

Table B.S1. Replicates of the parental isolates.....54

Table B.S2. Counts of A1 and A2 mating types among non-clone-corrected and clone corrected isolates with respect to year.....55

Table B.S3. Regions of differentiation between field F_1 and *in vitro* F_1 isolates associated with incidence of Mendelian errors.....56

Table B.S4. Genotype and haplotype counts for regions of differentiation between the field F_1 and *in vitro* F_1.....57

Table B.S5. Regions of differentiation between the field F_1 and inbred subpopulations.....58

Table B.S6. Genotype and haplotype counts in ROI-1 in the field population with respect to year and subpopulation.....59

Table B.S7. Genotype and haplotype counts in ROI-2 in the field population with respect to year and subpopulation.....60

Table B.S8. Regions of differentiation between isolates of opposite mating types in the field F_1 and inbred subpopulations. 61

CHAPTER 1

A (BRIEF) HISTORICAL PERSPECTIVE: OOMYCETES ARE NOT FUNGI

Bats are not birds
Dolphins are not fish
Oomycetes are not fungi

-Oomycete Molecular Genetics Network (OMGN)

Despite the emergence of cytological and genetic evidence in the early 1970s for *Phytophthora* as diploid (Sansome & Brasier, 1973), the scientific community did not coalesce around a taxonomical reclassification of oomycetes as distinct from true fungi until the 1990s. At this juncture, biochemical evidence and ribosomal RNA sequencing placed *Phytophthora* on indisputable taxonomical footing in the *Straminipilia* kingdom, with diatoms and brown algae (Gunderson *et al.*, 1987; Judelson & Blanco, 2005; Jiang & Tyler, 2012).

Indeed, correspondence among oomycete researchers reveals a bitter battle. Already in 1973, Eva Sansome introduced her second instantiation of the diploid state of *Phytophthora*: "In the midst of some controversy, cytological and genetical evidence is accumulating to support the view that, in contrast to other fungi, the Oomycetes resemble higher organisms in being diploid…" (Sansome & Brasier, 1973). Almost twenty years later, in 1992, mycologist Dr. Donald J. S. Barr stated unapologetically, "For some time I have been saying that Oomycetes are related to the heterokont algae without qualifications such as 'perhaps' or 'maybe' " (Barr, 1992). Similarly, Dr. David Moore of the University of Manchester, histrionically derided the

authors of *The Growing Fungus* (1995), Neil A. R. Gow and Geoffrey M. Gadd, for too amply treating oomycetes in a fungal text: "*Neurospora* is probably more closely related to a cow's nose than it is to *Saprolegnia*" (Moore, 1997).

In spite of the *Straminipilia* consensus, mycologists (and the emerging oomycetologists) were not entirely appeased. The polyphyletic nature of the fungal kingdom was still at stake! Even Barr was not immune to the emergent semantics dispute: "Yes, Oomycetes and chytrids are fungi." Wherein, Barr redefined fungi as the "Union of Fungi", an example of one such "…polyphyletic assemblages that include kingdoms, or parts of kingdoms, that make up logical groups for the benefit of the applied biologist, mycologist or phytopathologist" (Barr, 1992). In his poignantly titled 1998 article 'Why oomycetes have not stopped being fungi', Miami University's Dr. Nicolas P. Money (self-referentially "…of the endangered breed of mycologists to have maintained an interest in oomycetes…") rallied with Barr: "While a phylogenetic chasm separates the oomycetes from other fungi, I shall argue that it is impractical to restrict the usage of the term fungus to those microorganisms that qualify as members of the Phylum Fungi" (Money, 1998). It seems Money and Barr feared that the dissipation of oomycete research would accompany relinquishment of the fungal umbrella.

Yet, in valorizing the contributions of oomycetes to general mycological knowledge, Money portended the tenuousness of this fear: "If justification for mycological research on oomycetes is necessary…It is also worth remembering that the entire field of plant pathology might be said to have originated with Berkeley's study of the oomycete *Phytophthora infestans*" (Money, 1998). Perhaps of even greater significance, Anton de Bary's experiments with *P. infestans* and potato tuber rot in the 1840s, to which Money refers, marked the first time that disease was attributed to a microorganism, ultimately ushering in the acceptance of Louis

Pasteur's germ theory in favor of spontaneous generation (Large, 1940). These early, immense contributions of *Phytophthora* to scientific knowledge presaged the contemporary oomycete researcher rallying cry: "Oomycetes are not fungi."

Both tides of taxonomical animosity (not so) long since abated, contemporary oomycete researchers (no longer an "endangered breed" nor resting on the laurels of the fungal kingdom) embrace the challenges and idiosyncrasies of *Phytophthora* biology. Advent of inexpensive sequencing techniques and novel molecular biology tools, and a burgeoning research community, have discredited the "gloomy view of the 1980s that oomycetes are a 'fungal geneticist's nightmare' (Shaw, 1983; Schornack *et al.*, 2009)" (Kamoun *et al.*, 2014). Concerted and diverse inquiry in *Phytophthora* comes at a time when many of the >120 species in the *Phytophthora* genus (Martin *et al.*, 2012; 2014) continue to plague agricultural production and natural ecosystems (Erwin & Ribeiro, 1996; Kamoun *et al.*, 2014). Furthermore, *Phytophthora* species provide unique models to investigate reproductive biology, pathogenesis, population biology, and host resistance.

In the present thesis, I have undertaken the study of the heterothallic *P. capsici*, the causal agent of Phytophthora blight. Since its discovery in 1922 by Leon H. Leonian at the New Mexico Agricultural Research station (Leonian, 1922), *P. capsici* has been credited with the destruction of vegetable crops globally (Hausbeck & Lamour, 2004; Granke *et al.*, 2012; Lamour *et al.*, 2012). In the last decade, flooding events have catalyzed novel infestations of Phytophthora blight in the northeast United States. In this region, widespread distribution of both mating types, A1 and A2, results in the establishment of sexually reproducing populations (Dunn *et al.*, 2010). As the sexual spores, oospores, remain viable in the soil for many years, understanding the long-term trajectory of these populations is essential to improving management practices.

Herein, I characterize the temporal genetic dynamics of a closed, biparental, experimental field population of *P. capsici* utilizing genome-wide, single-nucleotide polymorphism (SNP) genotyping.

REFERENCES

Barr D, 1992. Evolution and kingdoms of organisms from the perspective of a mycologist. *Mycologia* **84**, 1.

Dunn AR, Milgroom MG, Meitz JC *et al.*, 2010. Population Structure and Resistance to Mefenoxam of *Phytophthora capsici* in New York State. *Plant Disease* **94**, 1461–1468.

Erwin DC, Ribeiro OK, 1996. *Phytophthora Diseases Worldwide*. St Paul, MN: The American Phytopathological Society.

Gow NAR, Gadd GM, 1995. *The Growing Fungus*. London, UK.

Granke LL, Quesada-Ocampo L, Lamour K, 2012. Advances in research on *Phytophthora capsici* on vegetable crops in the United States. *Plant Disease* **96**, 1588–1600.

Gunderson JH, Elwood H, Ingold A, Kindle K, Sogin ML, 1987. Phylogenetic relationships between chlorophytes, chrysophytes, and oomycetes. *Proceedings of the National Academy of Sciences* **84**, 5823–5827.

Hausbeck MK, Lamour KH, 2004. *Phytophthora capsici* on vegetable crops: research progress and management challenges. *Plant Disease* **88**, 1292–1303.

Jiang RHY, Tyler BM, 2012. Mechanisms and Evolution of Virulence in Oomycetes. *Annual Review of Phytopathology* **50**, 295–318.

Judelson HS, Blanco FA, 2005. The spores of Phytophthora: weapons of the plant destroyer. *Nature Reviews Microbiology* **3**, 47–58.

Kamoun S, Furzer O, Jones JDG *et al.*, 2014. The Top 10 oomycete pathogens in molecular plant pathology. *Molecular Plant Pathology* **16**, 413–434.

Lamour KH, Mudge J, Gobena D *et al.*, 2012. Genome Sequencing and Mapping Reveal Loss of Heterozygosity as a Mechanism for Rapid Adaptation in the Vegetable Pathogen *Phytophthora capsici*. *Molecular Plant-Microbe Interactions* **25**, 1350–1360.

Large EC, 1940. The advance of the fungi. *The Advance of the Fungi*.

Leonian LH, 1922. Stem and fruit blight of peppers caused by Phytophthora capsici sp. nov. *Phytopathology*.

Martin FN, Abad ZG, Balci Y, Ivors K, 2012. Identification and detection of Phytophthora: reviewing our progress, identifying our needs. *Plant Disease* **96**, 1080–1103.

Martin FN, Blair JE, Coffey MD, 2014. A combined mitochondrial and nuclear multilocus phylogeny of the genus Phytophthora. *Fungal Genetics and Biology* **66**, 19–32.

Money NP, 1998. Why oomycetes have not stopped being fungi. *Mycological Research* **102**, 767–768.

Moore D, 1997. Book Review. *The Mycologist* **11**, 140.

Sansome E, Brasier CM, 1973. Diploidy and Chromosomal Structural Hybridity in Phytophthora infestans. *Nature* **241**, 344–345.

Schornack S, Huitema E, Cano LM *et al.*, 2009. Ten things to know about oomycete effectors. *Molecular Plant Pathology* **10**, 795–803.

Shaw DS, 1983. *The perenosporales: a fungal geneticist's nightmare. In: Zoosporic Plant Pathogens (Buczacki, S.T., ed.)*. London: Academic Press.

CHAPTER 2

TEMPORAL GENETIC DYNAMICS OF AN EXPERIMENTAL, BIPARENTAL FIELD POPULATION OF *PHYTOPHTHORA CAPSICI*

Introduction

Phytophthora capsici is the filamentous, soil-borne oomycete plant pathogen responsible for Phytophthora blight, a disease inflicting significant annual crops losses worldwide (Erwin & Ribeiro, 1996; Hausbeck & Lamour, 2004; Granke *et al.*, 2012; Lamour *et al.*, 2012). Success of *P. capsici* is facilitated by its widespread ability to overcome fungicides (Lamour & Hausbeck, 2000), dearth of resistant cultivars (Granke et al., 2012), and large, diverse host range (comprising >15 plant families), including widely grown, economically important vegetable crops in the *Cucurbitaceae*, *Solanaceae*, and *Fabaceae* plant families (Satour & Butler, 1967; Hausbeck & Lamour, 2004; Tian & Babadoost, 2004). Extreme weather events often initiate new infestations by introducing inoculum into agricultural fields via flood waters (Dunn et al., 2010). Contaminated soil and infected plant material are commonly implicated in pathogen spread (Granke et al., 2012), however, *P. capsici* is not aerially dispersed (Granke *et al.*, 2009).

Once introduced into a field, the explosive asexual cycle of *P. capsici* catalyzes the rapid escalation of disease within a growing season. When exposed to water saturated conditions, a single sporangium can release 20-40 zoospores, each capable of inciting root, crown, or fruit rot, the characteristic symptoms of Phytophthora blight (Hausbeck & Lamour, 2004). For sexual reproduction, the heterothallic *P. capsici* requires two mating types, classically referred to as A1 and A2 (Erwin & Ribeiro, 1996). Exposure to mating type specific hormones (α1 and α2) stimulates production of the gametangia, subsequent outcrossing, and formation of recombinant

oospores (Ko, 1988). However, both mating types are hermaphroditic, and thus capable of self-fertilization (Shattock, 1986; Ko, 1988), which is thought to occur at a lower rate relative to outcrossing in *P. capsici* (Uchida & Aragaki, 1980; Dunn *et al.*, 2014).

While the asexual reproductive cycle directly inflicts crop damage, sexual reproduction confers several epidemiological advantages. First, unlike asexual propagules, oospores survive exposure to cold temperatures (Hausbeck & Lamour, 2004; Babadoost & Pavon, 2013). Thus, in regions with cold winter conditions, oospores are the primary source of overwintering inoculum (Bowers, 1990; Lamour & Hausbeck, 2003; Granke *et al.*, 2012). Second, oospores remain in the soil for years regardless of host availability, enabling the persistence of the pathogen between growing seasons and rendering eradication unfeasible. In the spring, in the presence of susceptible hosts, germinating oospores, potentially formed in distinct years, initiate the repeating, asexual reproductive cycle (Hausbeck & Lamour, 2004; Granke *et al.*, 2012).

Where both mating types coexist, sexual reproduction is associated with persistent pathogen populations, genetic diversity, and an approximate 1:1 ratio of A1 to A2 mating types (Lamour & Hausbeck, 2001; Dunn *et al.*, 2010). While asexual reproduction can increase the prevalence of a specific genotype within a sexually reproducing population, the inability of asexual propagules to survive cold winters (Hausbeck & Lamour, 2004; Babadoost & Pavon, 2013) implies that each year meiosis disrupts linkage between the particular combination of alleles observed within a clone (Kondrashov, 1988). As a consequence, sexual reproduction mediates the effects of clonal propagation on *P. capsici* population structure (Lamour & Hausbeck, 2001). Furthermore, in geographic regions where sexual reproduction occurs, genetic differentiation between field populations, even within close proximity, suggests that after an

initial introduction limited gene flow occurs between fields (Lamour & Hausbeck, 2001; Dunn *et al.*, 2010), consistent with a lack of aerial dispersal (Granke et al., 2009).

Given this infection scenario, i.e. an initial inoculation but no subsequent introductions, we would expect *P. capsici* populations to exhibit the signatures of a bottleneck event: reductions in genetic diversity and an increase in inbreeding over time, proportional to the number of founding isolates (Kirkpatrick & Jarne, 2000). (We define inbreeding strictly as inter-mating between related isolates, and reserve selfing to refer to self-fertilization events.) In populations which undergo a so-called founder effect, inbreeding is expected to decrease mean population fitness over time due to the expression of recessive deleterious alleles, i.e. the genetic load, in the homozygous state (Charlesworth & Charlesworth, 1987; Hartl & Clark, 2007). A related phenomenon, inbreeding depression, i.e. the difference in fitness between selfed and outcrossed progeny in a population (Kirkpatrick & Jarne, 2000), is considered a major driver of obligate outcrossing, and may contribute to maintenance of self-incompatibility in hermaphroditic plant species (Charlesworth & Charlesworth, 1987). Charting the genetic trajectory of isolated populations of *P. capsici* in the context of these processes, is essential to understanding pathogen evolution in an agriculturally relevant scenario.

Thus, in 2008, to investigate the response of *P. capsici* to a severe bottleneck, we established a closed, biparental field population, by inoculating a research field once with two heterozygous strains of opposite mating types. In a preliminary study, we tracked the allele and genotypic frequencies of five microsatellite markers in the field population from 2009-12 (Dunn et al., 2014). We demonstrated that sexual reproduction resulted in high genotypic diversity, a function of the proportion of unique isolates (Grünwald *et al.*, 2003), in 2009-11, with a

reduction in genotypic diversity in 2012. However, five markers afforded limited power to characterize population and individual level phenomena.

Therefore, in the present study, we analyzed isolates collected in 2009-13 from the *P. capsici* field population with genotyping-by-sequencing (GBS), a multiplexed reduced-representation sequencing technique, which assays single nucleotide polymorphism (SNP) markers distributed throughout the genome (Elshire *et al.*, 2011). The closed experimental field design excluded introduction of new alleles via migration, providing a unique opportunity to address the influence of inbreeding on population genetic phenomena in *P. capsici*. In high-density SNP genotyping isolates from the biparental field population, our goal was threefold: 1) Evaluate the effects of oospore survival on population structure; 2) Quantify the genome-wide incidence of inbreeding; and 3) Identify whether specific regions deviate from the rest of the genome in terms of changes in allele frequency.

Results

GBS of the experimental biparental isolates and validation. We genotyped 232 isolates collected from a closed, biparental field population of *P. capsici* from 2009-13. All field isolates were collected from infected plant tissue, and are therefore, by definition, pathogenic. Additionally, we genotyped 46 single-oospore progeny from an *in vitro* cross between the same founding parents as a reference for the field isolates for which generation was *a priori* unknown. Three of the *in vitro* progeny were identified as putative selfs by Dunn et al. (2014), which was confirmed by our analysis, and are hereafter referred to as *in vitro* selfs to distinguish them from the *in vitro* F_1 progeny. The A1 (isolate: 0664-1) and A2 (isolate: 06180-4) founding parents were genotyped 14 and 11 times, respectively, to estimate laboratory and genotyping errors (Table B.S1).

Out of the 401,035 unfiltered variant calls, initial site filters reduced the data set to 23,485 high-quality SNPs (Figure C.S1), with an average SNP call rate (i.e. the percentage of individuals successfully genotyped at each SNP) of 95.93% (median of 97.64%). The 23,485 SNPs were equally distributed among 307 scaffolds (scaffold size and number of SNPs were highly correlated (r^2=0.95), with an average SNP density of approximately 1 SNP every 2.5 kb. There was essentially no correlation between mean individual read depth and heterozygosity per SNP among all isolates (r^2=0.009, *P*-value=0.10), indicating that heterozygous calling post-filtering was robust to differences in mean individual sequencing coverage (Figure C.S2 A).

To assess genotyping accuracy, we compared biological and technical replicates of the parental isolates. Replicates of the A1 parent (n=14) and A2 parent (n=11), representing 4-5 distinct serial cultures, shared on average 98.30% (s=0.45%) and 98.17% (s=0.51%) alleles identity-by-state (IBS), respectively (Table B.S1). This corresponded to 3.60% (s=0.86%) discordant sites on average among non-missing genotypes between replicates. Lower average discordance ($\bar{x}$=2.86%, s=0.33%) between only replicates of the same parental culture (n=54 pairwise comparisons) suggested variation associated with distinct culture time points. Therefore, our overall genotyping error rate, inclusive of variation in mycelial and DNA extractions, but not different culture time points, was approximately 3%. Among technical replicates [same DNA sample (n=4) sequenced 3-4 times] the error rate was on average 2.95%, indicating that the majority of genotype discrepancies were attributed to sequencing and genotyping errors rather than distinct mycelial harvests. When we excluded heterozygous calls in each pairwise comparison (n=21) of the technical replicates, less than 0.0001% sites were discordant, indicating that heterozygote genotype discrepancies drove genotyping errors. As in

the total data set, the association between individual sequencing coverage and heterozygosity was negligible in both sets of parental replicates (Figure C.S2 B).

Phytophthora capsici reproduces asexually, therefore, it was theoretically possible to sample the same genotype from the field multiple times within a year. To remove the bias imparted on population genetic analyses by including clones, we retained only one isolate for each identified unique genotype (Milgroom, 1996). Pairwise identity-by-state (IBS) between replicates of the A1 and A2 parental isolates were compared to establish a maximum genetic similarity threshold to define clones (see Methods), akin to (Rogstad *et al.*, 2002; Meirmans & van Tienderen, 2004). Applying this threshold, we identified 160 unique field isolates out of the initial 232 field isolates (Table B.S2). Two *in vitro* isolates and one field isolate were identified as outliers with respect to deviation from the expected 1:1 ratio of allele depths at heterozygous sites (n=2) or heterozygosity (n=1), and subsequently removed (Figure C.S3; see Methods). Previous studies have shown that deviation from a 1:1 ratio of allele depths at heterozygous sites, the expectation for diploid individuals, is correlated with ploidy variation (Yoshida *et al.*, 2013; Rosenblum *et al.*, 2013; Li *et al.*, 2015), therefore these two isolates provide preliminary evidence for ploidy variation in *P. capsici*. After outlier removal, the final data set consisted of 159 field isolates, 41 *in vitro* F_1, and three *in vitro* selfs.

Clones did not appear in multiple years, consistent with the inability of asexual propagules to survive the winter (Hausbeck & Lamour, 2004; Babadoost & Pavon, 2013). After clone-correction, the A2 mating type was more represented in the field (A1:A2=65:94; $\chi 2$ test, P-value=0.02), a phenomenon also observed in the *in vitro* F_1 (A1:A2=16:25; $\chi 2$ test, P-value=0.16; Figure 2.1). The only exception was 2012, which may be explained by a smaller sample size in this year, artificially compounded by loss of several unique isolates (based on

microsatellite profiles (Dunn et al., 2014) in culture prior to this study. We observed lower genotypic diversity in 2012-13 (Table B.S2), consistent with Dunn et al. (2014).

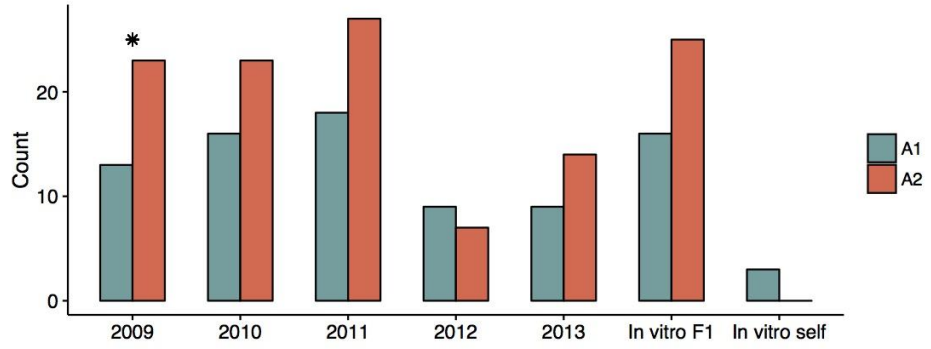


Figure 2.1. Distribution of the mating type of each isolate by year in the final, clone-corrected data set. Counts of the mating type of each isolate, A1 (teal) and A2 (reddish brown), in the *in vitro* F_1, *in vitro* selfs, and clone-corrected field isolates, separated by year. The star indicates a significant difference ($\chi 2$ test; *P*-value<0.1) between A1 and A2 counts.

To reduce oversampling of specific genomic regions, which can disproportionately influence population genetic inference (Price *et al.*, 2006; Abdellaoui *et al.*, 2013), without making assumptions about linkage disequilbrium (LD), we randomly selected one SNP within a given, non-overlapping 1 kb window. With final quality filters, and including only SNPs in scaffolds containing at least 300 kb (*n*=63), pruning resulted in a data set of 6,916 SNPs (Figure C.S1). Bimodal heterozygosity and minor allele frequency (MAF) distributions in this reduced SNP set were consistent with distributions in the unpruned data set (Figure C.S4). The pruned data set had a median SNP call rate of 98.01% and median site depth of 18.61 (i.e. average number of reads per individual per SNP). The median sample call rate (i.e. percentage of SNPs genotyped in each sample) was 97.77%, and the median sample depth (i.e. average number of reads per SNP per individual) was 20.36. Among technical replicates (*n*=4) the error rate was on average 1.52%. We utilized the pruned data set for all subsequent analyses.

Population differentiation increases with year. To broadly define genetic relationships between the *in vitro* and field isolates relative to the founding parents, we analyzed the field, *in vitro* and parents (represented by consensus parental genotypes, see Methods) jointly, with principal component analysis (PCA). The PCA exhibited the expected biparental population structure, in that the majority of isolates clustered in between the parental isolates along the major axis of variation, principal component (PC) 1 (Figure 2.2 A). Most 2009-11 isolates clustered with the *in vitro* F_1, whereas, many 2012-13 isolates were dispersed along both axes, suggesting differentiation associated with year.

To explore structure exclusively within the field population, we performed PCA on only the field isolates. Along PC1, isolates from 2012-13 were differentiated from prior year isolates (Figure 2.2 B). Whereas, PC2 described differentiation within and between years.

To assess the variance in allele frequencies between years, we estimated pairwise F_{ST} (Weir & Cockerham, 1984) between years, where each year was defined as a distinct population. All pairwise comparisons were significantly greater than zero with the exception of 2009 versus 2010. Small F_{ST} estimates for comparisons between 2009, 2010, 2011 and the *in vitro* F_1 indicated minimal variation in allele frequencies between these years. The greatest differences were observed between years 2012 and 2013 compared to 2009, 2010 and the F_1 populations (Figure 2.2 C), consistent with the PCA results. In addition, years 2012 and 2013 were also significantly differentiated from each other (F_{ST}=0.027).

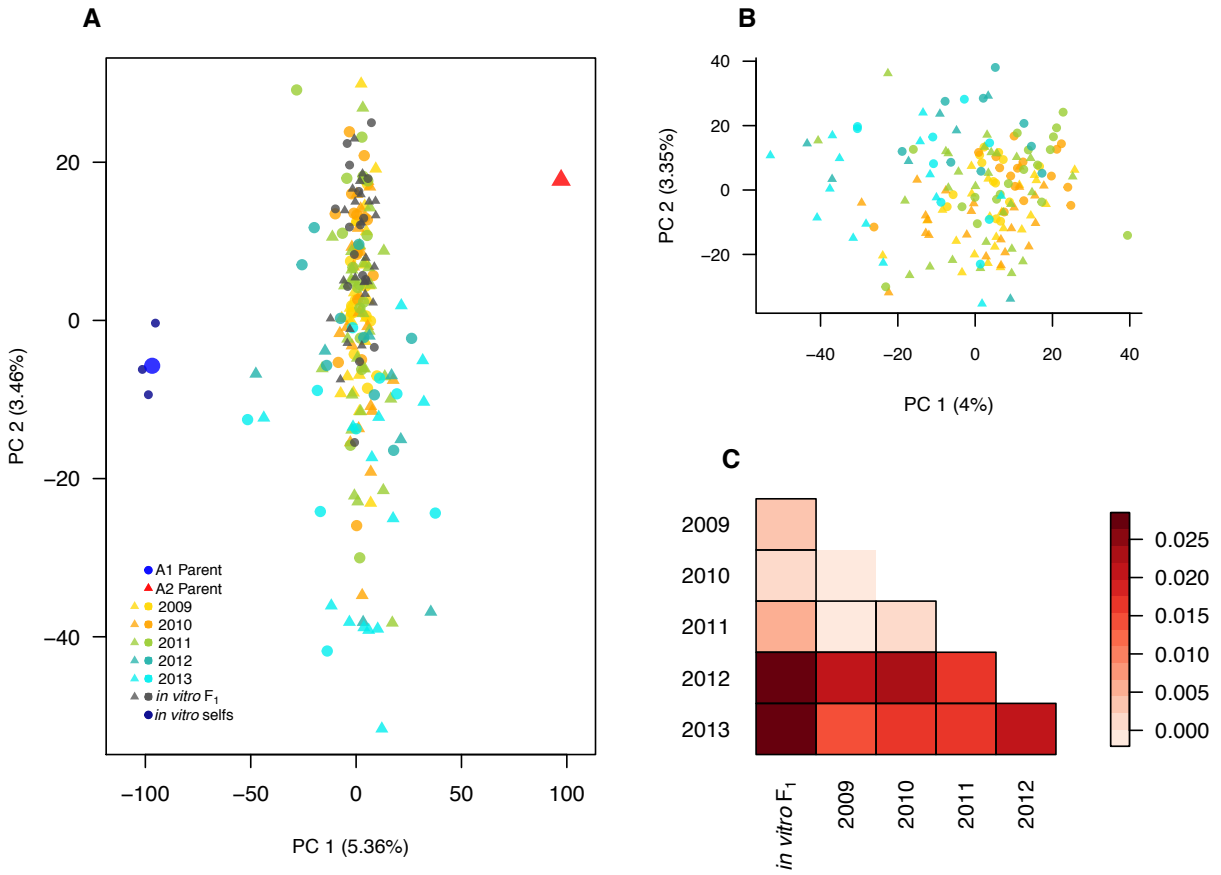


Figure 2.2. Population structure in the biparental field population relative to the *in vitro* F_1 and founding parents. A) Field isolates, *in vitro* F_1, *in vitro* selfs, and consensus parental genotypes plotted along the first two principal components (PCs). Each year is represented by a different color, with the A1 and A2 parental isolates indicated by blue and red, respectively. Shapes indicate the mating type of each isolate, with triangles (A1) and circles (A2). B) A PCA of only the field isolates, with color and symbol scheme consistent with (A). C) Pairwise F_{ST} for comparisons between sample years and the *in vitro* F_1 represented by a heat map, with more positive F_{ST} values increasingly red. A border indicates that the pairwise F_{ST} value was significantly different from 0, as tested by 1000 random SNP permutations.

Inbreeding in the field population. To quantify changes in inbreeding in the closed, field population, we estimated the individual inbreeding coefficient (F) for each isolate. While F does not directly measure identity-by-descent (IBD), it is highly correlated with IBD estimates in empirical and simulated data sets with relatively large numbers of markers (Kardos *et al.*, 2015), particularly in highly subdivided, small populations (Balloux *et al.*, 2004), such as the population under study. And, in a closed, biparental population, heterozygosity is directly proportional to the degree of inbreeding (Wright, 1921). Negative F estimates correspond to heterozygote excess relative to Hardy-Weinberg expectations for a reference population, defined here as the *in vitro* F_1. Positive F values indicate heterozygote deficiency.

First, to establish expectations for a known F_1, we assessed the F distribution in the *in vitro* F_1. The *in vitro* F_1, with a mean F of -0.366, was more heterozygous than the founding parents (average F across replicates=-0.007 (A1) and -0.183 (A2); Figure 2.3 A). In contrast to the unimodal *in vitro* F_1, the field population had a bimodal F distribution, with one peak approximately centered at the *in vitro* F_1 mean, and a second peak centered at a less negative F value. This second peak indicated that inbreeding was occurring in the field population.

To dissect the bimodal shape of the field distribution, we analyzed F for each year separately. Both for 2009 and 2010, the distributions were unimodal with F means not significantly different from the *in vitro* F_1 mean (pairwise t-test; *P*-values=1.0; Figure 2.3 B). For years 2012 and 2013, distributions were also unimodal, but had F means significantly less negative than the *in vitro* F_1 (*P*-values<0.0001). Year 2011 had a bimodal F distribution.

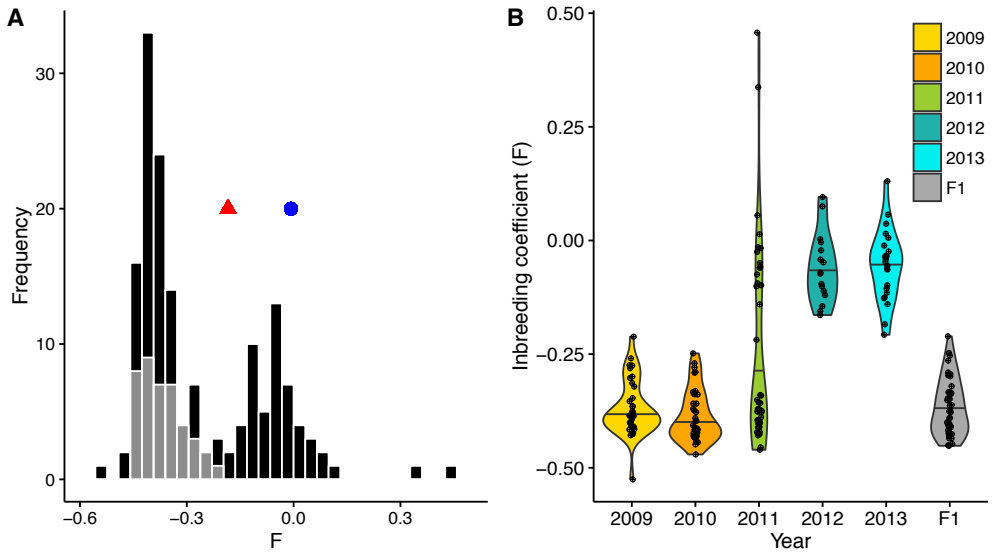


Figure 2.3. Generational shift in the field population. A) Superimposed histograms of the individual inbreeding coefficient (F), estimated from 6,916 SNPs, in the *in vitro* F_1 (gray) and field population (black). The *in vitro* F_1 were more heterozygous than the founding parental isolates, corresponding to negative F values, indicated by a blue circle (A1 parent) and red triangle (A2 parent). In contrast, the field population exhibited a bimodal F distribution. B) Distributions of F by year represented by violin plots, with each year represented by a distinct color and individual data points overlaid. The long upper tail of the 2011 distribution is driven by two field selfs.

To interpret the effect of changes in inbreeding on genotypic and allele frequencies with time, we analyzed both SNP heterozygosity and MAF distributions for each year. In a biparental cross, clear expectations for these quantities in the F_1 generation makes them informative in distinguishing F_1 from inbred generations. Specifically, in the F_1 generation, sites should segregate with a MAF of either 0.25 (for a cross of *Aa* x *AA*) or 0.5 (for *Aa* x *Aa* and *AA* x *aa*), and population heterozygosity should be 50% or 100% at each SNP. In the F_2 generation, i.e. a population derived from a single generation of inbreeding, MAF should remain constant, whereas heterozygosity should decline. Our results showed that the *in vitro* F_1, 2009, and 2010

behaved in accordance with expectations for a predicted F_1; the heterozygosity distributions had peaks centered at approximately 50% and 100% (Figure 2.4 A), and the MAF distributions had peaks at 0.25 and 0.5 (Figure 2.4 B). In contrast, MAF and heterozygosity distributions in 2012 and 2013 were not consistent with F_1 expectations, in that we no longer observed obvious peaks (Figure 2.4). While genotypic frequency shifts in 2012 and 2013 indicated presence of inbreeding and deviation from F_1 expectations, changes in the MAF distribution also denoted that these were likely not canonical F_2 populations. Discrete generations are implicit in a F_2, therefore deviation from F_2 expectations may be attributed to violation of this assumption.

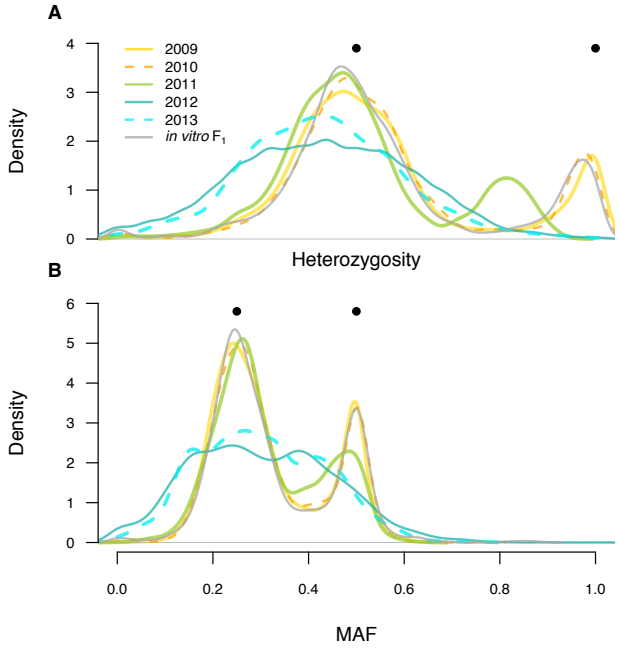


Figure 2.4. Year heterozygosity and allele frequency (MAF) distributions. Filled, black circles indicate expectations for population heterozygosity and MAF in a theoretical F_1 population. A) Distributions of the proportion of heterozygous individuals per SNP (n=6,916) for each year and the *in vitro* F_1, represented by kernel density estimates, with color corresponding to year. Bimodal distributions in the *in vitro* F_1 and years 2009-10 are consistent with expectations for the F_1 generation, whereas unimodal distributions in 2012-13 indicate presence of inbreeding. A shift in the bimodal distribution of 2011, indicates the mixed outbred and inbred composition of this year. B) MAF distributions, where the minor allele is defined based on the frequency in the total field population, for each year and the *in vitro* F_1, with color designations the same as in (A).

Finally, in 2011, both heterozygosity and MAF distributions were bimodal, as in an F_1, but with reduced heterozygosity and deviation in allele frequencies relative to the *in vitro* F_1 and prior years (Figure 2.4). These shifts in 2011 suggested coexistence of both F_1 and inbred isolates (i.e. non-F_1 isolates) in this year, consistent with the bimodal 2011 F distribution (Figure 2.3 B).

Selfing in the laboratory and field. In addition to quantifying inbreeding (defined as inter-mating between related isolates), we also estimated the incidence of self-fertilization in the biparental, field population. The frequency at which *P. capsici* reproduces through self-fertilization in either field or lab conditions is unknown (Dunn et al., 2014). Given the limited prior evidence of selfing in *P. capsici*, we first confirmed that the three putative *in vitro* selfs were indeed the product of self-fertilization by the A1 parent, as hypothesized by Dunn et al. (2014). To this end, we distinguished the *in vitro* selfs from the *in vitro* F_1 by four features: 1) Clustered with the A1 parent in PCA (dark blue circles in Figure 2.2 A); 2) Alleles shared IBS disproportionately with the A1 versus A2 parent; 3) Heterozygosity approximately 50% of the A1 parent; and 4) Significantly higher inbreeding coefficients relative to the F_1 (>3*s* from the mean; Table 2.1).

	Consensus parental genotypes		*in vitro* selfs			Putative field selfs	
Statistic	A1	A2	68_14	68_19	68_27	11PF_21A	11PF_26A
Individual heterozygosity[1]	0.41	0.48	0.25	0.21	0.22	0.22	0.27
F[2]	-0.01	-0.18	0.38	0.48	0.46	0.46	0.34
MEs[3]	0.19	0.19	0.25	0.28	0.27	0.27	0.21
IBS with the A1 parent[4]	1.00	0.47	0.91	0.89	0.90	0.65	0.66
IBS with the A2 parent[4]	0.47	1.00	0.45	0.44	0.44	0.62	0.65

[1] Proportion of heterozygote, non-missing sites per individual

[2] Individual inbreeding coefficient (F)

[3] Proportion of Mendelian errors (MEs) per individual (corrected for outlier SNPs), relative to the consensus parental

[4] Identity-by-state (IBS) of each isolate with the A1 or A2 parental consensus genotype

Table 2.1. Selfing *in vitro* and in the field. Characterizing selfed isolates in the *in vitro* and bi-parental field populations in terms of heterozygosity, Mendelian errors (MEs), and alleles shared identity-by-state (IBS) with either founding parent

Having shown that generalized expectations for selfing applied to *P. capsici*, we utilized extreme heterozygote deficiency as an indicator of selfing in the field. As, in the field context, the first three aforementioned selfing features were inapplicable because the progenitor of a selfed isolate in the field was not *a priori* known. We observed that two of the 2011 field isolates were F outliers (>3*s* from the mean) with respect to the inbred field contingent distribution ($\bar{x}$=-0.050, *s*=0.12). We classified these two A1 field isolates as field selfs (Table 2.1). Lack of disproportionate IBS of the field selfs with either founding parent denoted that these isolates were not the product of self-fertilization by either founding parent. Therefore, we observed selfing in the *in vitro* and field populations at frequencies of 3/46 (6.5%) and 2/159 (1.26%), respectively, denoting minimal incidence of selfing in both lab and field scenarios.

Classifying F_1 versus inbred isolates in the field using Mendelian errors (MEs).

Based on the above results, we hypothesized that 2009-10 were comprised of mainly F_1, 2012-13 inbred, and 2011 a mixture of both F_1 and inbred isolates. However, we had heretofore not verified that each year was homogeneous with respect to F_1 and inbred composition. To quantify the number of F_1 isolates, we used the fact that the genotypes of the founding parents were known to calculate an additional individual summary statistic, the proportion of Mendelian errors (MEs). A ME is defined as a genotype inconsistent with the individual being an F_1 derived from specific parents (Purcell et al., 2007), here, the A1 and A2 founders. Commonly, MEs have been used to detect genotyping and experimental errors in SNP data sets where pedigree information is known (Purcell et al., 2007). The expectation is that a true F_1 individual should have very few MEs, a postulate we applied to assess whether each field isolate belonged to the F_1 generation.

Initial ME estimates revealed both randomly distributed and clustered ME-enriched SNPs. In Appendix A.1, we show that clustered ME-enriched SNPs corresponded to inferred mitotic LOH events in the parental isolates in culture. After removing all ME-enriched SNPs (n=848), mean MEs per isolate for the *in vitro* F_1 and field F_1 subpopulations were 1.38% and 0.98%, below our estimated genotyping error rate of approximately 1.5%.

Akin to F, the proportion MEs per individual is a function of genotypic frequencies. Therefore, it was not surprising that year distributions of the ME statistic were consistent with F, with increased MEs in years 2012-13 (Figure 2.5 A). Because asexual propagules do not survive the winter (Hausbeck & Lamour, 2004; Babadoost & Pavon, 2013), it can be assumed that all F_1 isolates in the field, in any year, were from oospores in the year of the initial field inoculation (2008). Applying a threshold of 5.58% MEs (3s from the *in vitro* F_1 mean) to characterize F_1 versus non-F_1, we observed exclusively F_1 in 2009-10, a mixture of F_1 and inbred isolates in

2011 (ratio of F_1 to inbred=29:16) and all inbred isolates in 2012-13 (Figure 2.5 B). As such, F_1 dominated in 2009-11, demonstrating that oospores were viable and pathogenic for at least three years.

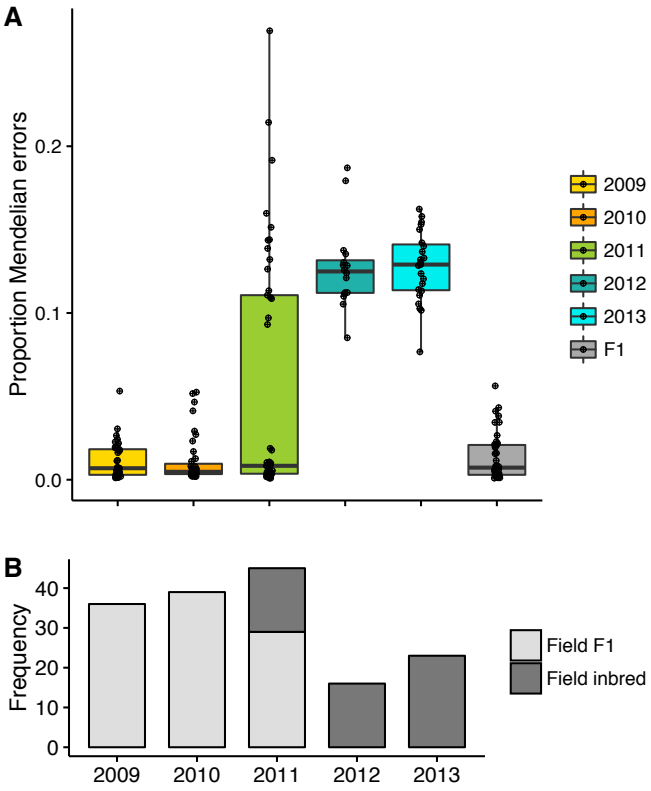


Figure 2.5. Mendelian errors (MEs) distinguish F_1 and inbred isolates in the field. A) Boxplots of the proportion of MEs per individual for each year are consistent with the inbreeding coefficient trend, with a bimodal distribution in 2011, and increased MEs in later years. B) Classification of each isolate based on the proportion of MEs, with counts of field F_1 (light gray) and field inbred (dark gray) for each sample year.

When the inbred isolates were removed from the 2011 data, the MAF distribution for 2011 was consistent with F_1 expectations (Figure C.S5 A). Concurrent observation of both F_1 and inbred isolates in a single year (2011) provided direct evidence of overlapping generations in

the field population, supporting overlapping generations as contributing to deviation from F_2 expectations in the inbred 2012 and 2013 years.

In addition, the ME estimates allowed us to pool isolates from separate years to define sub-populations, the field F_1 (*n*=104) and the field inbred (*n*=53; excluding the field selfs), for subsequent analyses. As in the total field population, A2 isolates were overrepresented in both the field F_1 and inbred subpopulations (A1:A2=43:61 and 21:32, respectively).

Regions of differentiation between generations in the field population. The generational transition in the field population from F_1 to inbred was accompanied by changes in the MAF distribution (Figure C.S5 B), implying the biased transmission of alleles to generations beyond the F_1. To identify which SNPs drove this allele frequency shift, we performed a genome-wide Fisher's Exact test of allele frequency differences between the field F_1 and field inbred subpopulations. We collectively analyzed these two subpopulations, rather than compare allele frequencies between years, due to the presence of overlapping generations, which complicate interpretation of temporal dynamics (Jorde & Ryman, 1995). From this analysis, we observed several regions of differentiation between these subpopulations (Figure 2.6 A; see Table B.S5 for coordinates).

First, we focused on the region with the most highly differentiated SNP, referred to as region of interest 1 (ROI-1; Figure 2.6 B). Of the 94 SNPs spanned by ROI-1, 44% were among SNPs in the top 2% of loadings for PC1 in the field PCA, showing that this region was correlated with differentiation in the field population. To assess the relationship between allele frequency changes and parental haplotype frequencies, we locally phased all isolates using a deterministic approach (see Methods). Haplotypes in ROI-1 (H1a, H3a, and H4a) were defined based on the sub-region (251,367-560,094 bp) which contained the majority of significantly differentiated

SNPs (44 out of 52 SNPs) and formed a LD block (Figure C.S15 A).

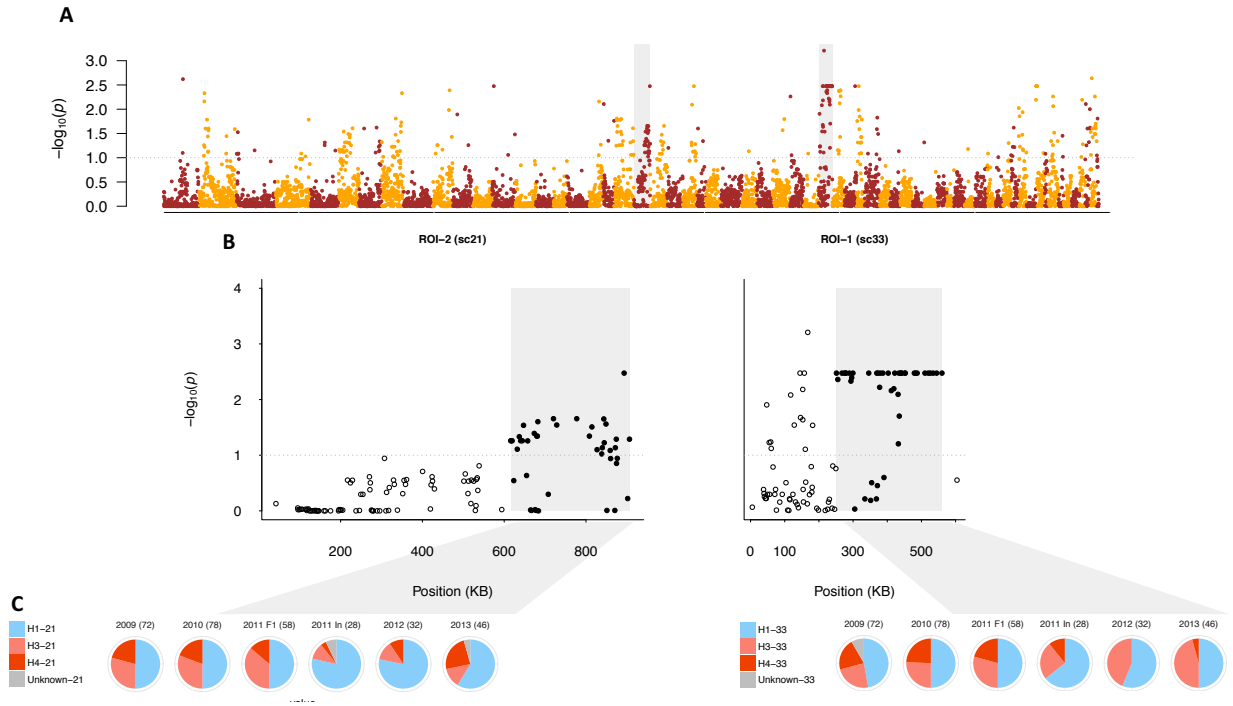


Figure 2.6. Regions of differentiation between the field F_1 and inbred subpopulations. A) Negative $\log_{10}$-transformed, false-discovery rate (FDR) adjusted *P*-values from the genome-wide test of allele frequency differences between the field F_1 and inbred subpopulations, ordered by physical position. The gray dotted lines in (A) and (B) indicate the significance threshold (α=0.10). Color alternates by scaffold. The shaded gray boxes indicate the SNPs in scaffolds 21 and 33 corresponding to ROIs 2 and 1, respectively. B) Same as (A) except that *P*-values are shown only for scaffolds 21 and 33. Here, gray boxes denote the sub-region within each scaffold defined as a ROI. Closed, black circles indicate SNPs within each ROI, whereas open, black circles indicate SNPs outside of the ROI. C) Pie charts represent the haplotype frequencies found in each year (with 2011 separated into F_1 and inbred (In) isolates), with the number of sampled chromosomes noted for each year. Blue corresponds to the single A1 founding parental haplotype, shades of red to the two A2 founding haplotypes, and gray to undesignated haplotypes in each ROI (see Methods).

Segregation among the F_1 isolates in each year (2009 to 2011) followed the F_1 expectation of a 2:1:1 ratio of H1a:H3a:H4a haplotypes (χ^2 test; *P*-values=0.91, 0.99, 0.65,

respectively). In contrast, in 2011 (inbred isolates only), 2012, and 2013, we observed lower frequencies of H4a and higher frequencies of H3a relative to the field F_1 subpopulation (Figure 2.6 C). The decline in H4a frequency from 22.12% in the field F_1 to 4.72% (and corresponding increases in H3a and H1a) in the field inbred drove allele frequency changes in ROI-1 (Table B.S6). Because the H4a sequence was most distinct from the other haplotypes, the reduction in H4a frequency, along with inbreeding, resulted in declines in heterozygosity in ROI-1. Consistent reductions in H4a frequency among inbred isolates in 2011-13 compared to F_1 isolates in prior years, provided strong evidence for the influence of selection. However, absence of H4a in year 2012 is very likely an artifact of smaller sample size in this year.

We next focused on a region in scaffold 21, defined as ROI-2, with the highest density of significantly differentiated SNPs (67%; Fig 2.6 B). In ROI-2, as in ROI-1, only three haplotypes segregated in the field population (Table B.S7). While not significant (at α=0.05), segregation among the F_1 isolates in each year (2009 to 2011) deviated from the F_1 expectation of a 2:1:1 ratio of H1:H3:H4 haplotypes (χ^2 test; *P*-values=0.61, 0.35, and 0.05, respectively), primarily attributed to higher H3 versus H4 haplotype frequency in the field F_1 (χ^2 test; *P*-value<0.01). A decline in frequency of the A2 parent haplotype, H3, by 19.47% and an increase in the A1 parent haplotype, H1, by 19.81% drove allele frequency changes (Fig 2.6 C; Table B.S7). While the frequency of H3 and H4 oscillated among inbred isolates in 2011-13, the H1 haplotype frequency was consistently higher than in the field F_1. In addition, we observed a high frequency of homozygous H1 genotypes (53%), whereas the H3 and H4 haplotypes were not observed in the homozygous state in the field inbred subpopulation, contrary to expectations (Table B.S7).

To posteriorly assess the significance of changes in allele frequency in ROIs 1 and 2, we compared the median F_{ST} value for significantly differentiated SNPs in each of these regions to

the genome-wide SNP F_{ST} distribution, where F_{ST} was defined as in (Lewontin & Krakauer, 1973). Assuming that drift acts equally throughout the genome, extreme deviations in F_{ST} provide evidence for selection (Lewontin & Krakauer, 1973). Median observed changes in allele frequency in ROIs 1 and 2 were in the in the 97th and 98th percentiles, respectively, with respect to genome-wide F_{ST}, showing that allele frequency changes in these regions vastly exceeded the genome-wide average.

Heterozygosity declines are slower in the mating type region. To investigate whether the mating system was a direct driver of differentiation in the field population, we first identified mating type associated SNPs using a Fisher's exact test of allele frequency differences between isolates of opposite mating types in the field F_1 (n_{A1}=43 and n_{A2}=61; Figure C.S15 B). The majority of the 184 significantly differentiated SNPs were in sub-regions of scaffolds 4 (37%) and 27 (43%), with additional differentiated SNPs in sub-regions of scaffolds 2, 34, and 40 (Fig 2.7 A; Table B.S7). All scaffolds containing significantly associated SNPs were in linkage group 10, consistent with a prior study (Lamour et al., 2012), and supporting presence of a single mating type determining region in *P. capsici*, as posited for *P. infestans* and *P. parasitica* (Fabritius & Judelson, 1997). SNPs in these five sub-regions comprised 20.29% of SNPs with elevated PC loadings (top 2%) in the PCA of only the field isolates, compared to 5.10% genome-wide, denoting that these SNPs were disproportionately correlated with differentiation in the field population.

At 98.30% of the *AA* x *Aa* SNPs associated with mating type in the field F_1, the A2 parent was heterozygous (*Aa*) and the A1 parent was homozygous (*AA*). As such, heterozygosity in the field progeny at these SNPs was attributed to inheritance of the minor allele (*a*), descendent originally from the A2 parent. Therefore, segregation of the A2 but not A1 parental haplotypes

was predominantly associated with mating type in the field F_1.

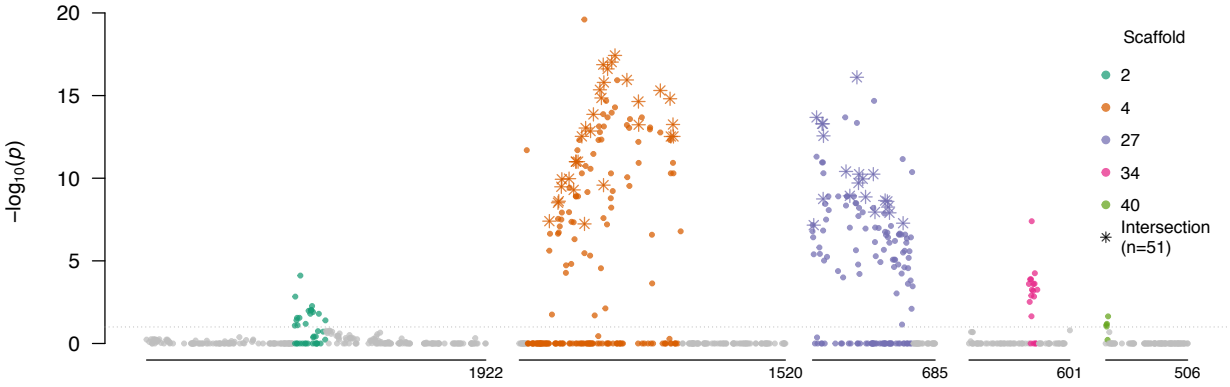


Figure 2.7. Allele frequency differences between isolates of opposite mating types. Negative $\log_{10}$-transformed *P*-values, adjusted for multiple testing, from the Fisher's exact test of allele frequency differences between A1 and A2 isolates in the field F_1, plotted against physical position, for scaffolds with significantly differentiated regions (Table B.S8). Colored SNPs were within the bounds of the minimum and maximum significant SNPs in each scaffold containing at least two significantly associated SNPs within 200 kb. Stars indicate the SNPs which were significant in tests of allele frequency differences between mating types in both the field F_1 and inbred subpopulations (Appendix A.3). All SNPs above the gray horizontal line were significant after the FDR correction (α=0.1).

We defined the mating type region (MTR) as consisting of genomic tracts encompassed by the minimum and maximum significant SNPs in scaffolds 4 and 27, which comprised 1.42 of the 1.64 Mb spanned by the five sub-regions, and contained 81% of the significantly differentiated SNPs. While we refer to a singular MTR, this was not intended to imply physical linkage between these two scaffolds. Based on the 293 SNPs in the MTR, the PCA of all isolates (*in vitro* and field; n=203) showed incomplete differentiation according to mating type (Figure C.S16).

To assess changes in heterozygosity in the MTR, we compared the heterozygosity distributions of the field F_1 (n_{A1}=43 and n_{A2}=61) and inbred (n_{A1}=21 and n_{A2}=32) isolates in the

MTR to the respective genome-wide distributions (see Methods). Observed heterozygosity in the field F_1 in the MTR was not centered at a significantly greater mean than the field F_1 genome-wide distribution (one-sided Wilcoxon rank-sum test, all *P*-values>0.87; Figure C.S17 A). In contrast, observed heterozygosity in the field inbreds in the MTR was shifted towards a greater mean relative to the field inbred genome-wide distribution (all *P*-values<0.005; Figure C.S17 A). Therefore, in the field inbred subpopulation, heterozygosity declines were less appreciable in the MTR compared to the rest of the genome. In addition, we found that heterozygosity in the MTR was significantly higher than the rest of the genome for both the A1 and A2 isolates in the field inbred subpopulation (Figure C.S17 D; all *P*-values<10^{-4} and 0.003, respectively), but not in the field F_1 (Figure C.S17 C; all *P*-values>0.99 and 0.65, respectively). Yet, heterozygosity in the MTR did not significantly exceed HWE expectations for A1 inbred isolates, as observed in the A2 inbred isolates, and both mating types in the field F_1 (Figure C.S18). These results were replicated when the A2 inbred isolates were down-sampled to the A1 inbred sample size (data not shown).

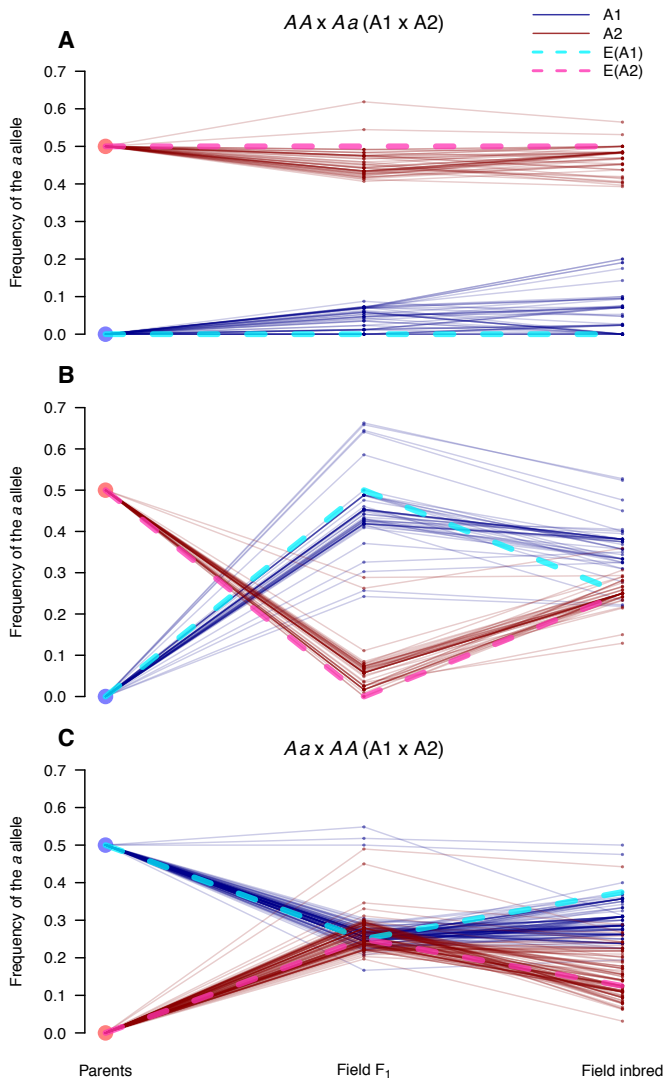


Figure 2.8. Segregation of SNPs in the mating type region follow expectations for sex-linked loci. Frequency of the *a* allele (p_a), for *AA* x *Aa* and *Aa* x *AA* (A1 x A2) markers in the mating type associated sub-regions of scaffolds 4 and 27, defined as the mating type region (MTR). Each parallel coordinate plot (A-C) tracks p_a at three time points (parents, field F_1, and field inbred) in the A1 (blue solid lines) and A2 (red solid lines) isolates for: A) *AA* x *Aa* markers (n=49) with p_a>0.3 in the A2 and p_a<0.3 in the A1 field F_1 isolates; B) Remaining *AA* x *Aa* markers (n=49); and C) All *Aa* x *AA* markers. Expectations for sex-linked loci, indicated by dotted lines, assuming that the A1 and A2 mating type behave like the homogametic (light blue) and heterogametic (pink) sexes, respectively, when: A) the *a* allele is in the A2 determining haplotype (i.e. Y analog); B) the *a* allele is in the non-A2 determining haplotype (i.e. X analog); and C) the *a* allele is in either of the non-A2 determining haplotypes (i.e. X analog).

To further dissect the genetic dynamics of mating type in the field population, we tracked the allele frequencies of markers that were heterozygous in one parent and homozygous in the other parent (*AA* x *Aa*). These markers are particularly informative because the origin of the *a* allele can be unambiguously assigned to the heterozygous parent. Specifically, we calculated the frequency of the parental tagging allele (p_a) in the parents, the field F_1, and the field inbreds at each of the *AA* x *Aa* SNPs in the mating type associated sub-regions of scaffolds 4 and 27 (n_{AAxAa}=206), for each mating type separately.

For A2 tagging SNPs (*n*=98), i.e. SNPs heterozygous in the A2 founding parent, we observed two classes of markers: those with p_a of approximately 0.5 in the A2 and 0.0 in the A1 field F_1 isolates, and the opposite scenario (Figure 2.8 A-B). In the first case (*n*=49), differences in p_a between mating types were maintained in the field inbred subpopulation (Figure 2.8 A). Whereas, in the second case (n=49), the difference in p_a between mating types narrowed (Figure 2.8 B). In contrast to the A2 parent tagging SNPs (*n*=108), markers heterozygous in the A1 parent and homozygous in the A2 parent predominantly followed a single pattern. These markers were at approximately equal allele frequencies in both mating types in the field F_1, but slightly diverged in frequency in the field inbred subpopulation. These three distinct segregation patterns were consistent with the association of presence/absence (P/A) of one of the A2 founding haplotypes in association with mating type.

While the structural basis of mating type determination in *P. capsici* is not known, observed segregation patterns in the mating type region resemble those of an XY system, where P/A of the Y determines sex. Therefore, as frame of reference, we derived expectations for sex-linked loci in an XY system, i.e. loci conserved between both sex chromosomes (Clark, 1988; Allendorf *et al.*, 1994), assuming that the A2 parent corresponded to the heterogametic sex.

Using this model, expectations (blue and pink dotted lines in Figure 2.8) closely matched the observed p_a trajectories in all three cases (A-C), further supporting the association of one of the A2 haplotypes with mating type determination.

Discussion

To study the temporal genetic dynamics of *P. capsici* in response to a severe bottleneck, we SNP genotyped at high-density 232 isolates collected in years 2009-13 from a closed, biparental field population founded in 2008, in Geneva, NY (Dunn et al., 2014). This experimental population parallels the infection scenario of a natural *P. capsici* epidemic, where a limited number of pathogen strains are thought to found a subsequently isolated population (Lamour *et al.*, 2012; Dunn *et al.*, 2014). Using GBS, we identified 159 unique field isolates and obtained 6,916 high quality SNPs with high sequencing depth (~20X coverage), low missing data, and over 97% reproducibility of genotype calls, distributed throughout the genome. With these data, we assessed temporal heterozygosity and allele frequency changes in the biparental population, representing the only controlled, multi-year genomic field study of a plant pathogen to date.

With knowledge of the parental genotypes and assuming simple Mendelian inheritance, we developed a threshold to detect F_1 field isolates based on the incidence of MEs in the *in vitro* F_1 progeny. Our results showed that both field and *in vitro* F_1 progeny were characterized by individual heterozygosity in large excess of Hardy-Weinberg expectations, explained by the fact these isolates were descendent from only two parents. With small numbers of parents, the probability of allele frequency differences between opposite sexes (here, mating types) increases,

consequently resulting in deviation from HWE among the progeny (Robertson, 1965; Pudovkin *et al.*, 1996; Luikart & Cornuet, 1999; Balloux, 2004).

Over time, the field population underwent a generational shift, transitioning from F_1 in 2009-10, to mixed generational in 2011, and ultimately all inbred in 2012-13. Presence of exclusively F_1 in 2009 suggests that the vast majority of oospores formed in the founding year (2008) were F_1. As oospores require a dormancy period of approximately one month (Satour & Butler, 1968; Dunn *et al.*, 2014), it is not surprising that there was insufficient time to produce multiple generations in the founding year. The presence of only F_1 and no inbred isolates in 2010, however, cannot be similarly explained. Rather, abundant sexual reproduction in the founding year, coupled with a lower rate in 2009, may have led to disproportionate presence of F_1 oospores (from 2008) surviving in the soil and germinating in 2010. Year 2011, where both inbred and F_1 isolates were observed in the field, signified a generational shift in the population. The absence of F_1 in the following years (2012-13) is consistent with previous reports of oospore declines in viability over time (Bowers, 1990), and negligible oospore survival after four years in field conditions (Babadoost & Pavon, 2013). While we did not quantify disease incidence in the field, observation of predominantly F_1 isolates in 2010-11 suggests that highly productive years contributed disproportionately to population structure, in accordance with theoretical predictions for populations in which sexual propagules require a dormancy period, e.g. plant species with seed banks (Templeton & Levin, 1979; Nunney, 2002). As a consequence, heterozygosity did not immediately decline in the second year of the field population, similarly consistent with the delayed attainment of equilibrium genotypic frequencies attributed to seed bank dynamics (Templeton & Levin, 1979).

Approximate equilibrium genotypic frequencies were not observed in the field population until the fourth year (2012). Here, a single large increase in homozygosity in the total population was consistent with cycles of inbreeding beyond a theoretical F_2 resulting in less appreciable declines in heterozygosity relative to the prior generation (Wright, 1921). However, two excessively homozygous field isolates, identified as field selfs, significantly deviated from this trend. Given this low frequency of selfing, we conclude that *P. capsici* behaved essentially as an obligate outcrossing species in the biparental field population. Occurrence of selfing in *P. capsici* is consistent with a previous report of oospore induction when strains of opposite mating types were separated by a membrane (Uchida & Aragaki, 1980), but contradicts previous studies which found no evidence for self-fertilization under *in vitro* conditions (Hurtado-Gonzales & Lamour, 2009; Lamour et al., 2012). As a single generation of self-fertilization reduces heterozygosity by approximately 50% in the progeny, minimal incidence of selfing delayed potential heterozygosity declines in the field population, as described for hermaphroditic plant species (Balloux, 2004).

In addition, while we observed three A1 parental selfs among the 46 *in vitro* progeny, we did not observe selfs derived from either founding parent in the field, despite the larger field F_1 sample size. Field isolates were inherently selected for both viability and pathogenicity, as well as resilience to environmental factors, whereas *in vitro* isolates were selected solely on viability in culture. Therefore, this result may reflect a fitness cost to self-fertilization, as observed in essentially all outcrossing species (Charlesworth & Charlesworth, 1987; Falconer & Mackay, 1996), manifest to a greater extent in the field versus laboratory conditions.

Given the potential fitness cost to self-fertilization in *P. capsici*, an increase in inbreeding may explain the allele frequency changes which accompanied the transition from an F_1 to inbred

population, as inbreeding presents recessive deleterious alleles in the homozygous state, rendering them subject to selection (Kirkpatrick & Jarne, 2000; Charlesworth, 2003). Simultaneously, inbreeding indirectly influences allele frequencies by decreasing the effective population size (N_e) relative to the census population size, consequently amplifying the effects of genetic drift (Charlesworth, 2009). In addition to inbreeding, many other factors likely decreased N_e, thereby increasing the influence of genetic drift: imbalanced sex (here, mating type) ratios (Charlesworth, 2009); clonal reproduction (Balloux *et al.*, 2003); variation in reproductive success (Hartl & Clark, 2007); small population sizes (Hartl & Clark, 2007) indicated by lower genotypic diversity in 2012-13; and overlapping generations (Felsenstein, 1971). Conversely, minimal differentiation between 2009-11, denotes that oospore survival, in behaving like a seed bank, mitigates the aforementioned reductions in N_e by maintaining a reservoir of genetic variation in the soil (Templeton & Levin, 1979; Hairston & De Stasio, 1988; Nunney, 2002; Waples, 2006).

In contrast to the general trends described above, we characterized two regions (ROI-1 and ROI-2) that significantly deviated from the genome-wide distribution of allele frequency differences (median allele frequency change in the 97 percentile or greater) between the field F_1 and inbred subpopulations. In these two regions, which presented only three segregating haplotypes, in contrast to the expected four, for heterozygous parents, we associated allele frequency changes with haplotype frequency shifts. Genetic drift may still explain these results, as drift has a larger effect in regions of low variation, i.e. with higher effective inbreeding coefficients corresponding to lower local N_e (Charlesworth, 2003; 2009). However, extreme changes in allele frequency are also suggestive of natural selection (Lewontin & Krakauer, 1973; Galtier *et al.*, 2000). Here, observation of corresponding haplotype frequency shifts is consistent

with hitchhiking or background selection having a large effect in inbred populations (Charlesworth, 2003), particularly with only a few generations. Alternatively, mitotic LOH, a phenomenon reported in the present study (see Appendix A.2) and in numerous *Phytophthora* species (Chamnanpunt *et al.*, 2001; Grünwald *et al.*, 2012; Lamour *et al.*, 2012; Kasuga *et al.*, 2016), may explain the observation of a disproportionate number of homozygous genotypes among inbred isolates in ROI-2. Evidence for mitotic LOH in numerous species, e.g. *Saccharomyces cerevisiae* (Magwene et al., 2011), *Candida albicans* (Forche et al., 2011), and the chytrid *Batrachochytrium dendrobatidis* (Rosenblum et al., 2013), supports the theoretical expectation that this process facilitates adaptation by interacting with selection to alter allele frequencies (Mandegar & Otto, 2007). Given the limited number of generations, we cannot unequivocally attribute these dramatic haplotype frequency shifts to selection. Furthermore, additional work is required to assess the role of these regions in pathogenicity and local adaptation.

While we observed a genome-wide increase in homozygosity in the field population due to inbreeding, reductions in heterozygosity in the identified mating type associated region were smaller relative to the genome for both A1 and A2 isolates. We show that this result is explained by persistent allele frequency differences between isolates of opposite mating types in the MTR. Maintenance of elevated heterozygosity in sex-linked regions has been attributed to differences in founding allele frequencies between sexes in several systems (Allendorf *et al.*, 1994; Marshall *et al.*, 2004; Waples, 2014). Further, *AA* x *Aa* SNPs associated with mating type in the field F_1 were predominantly heterozygous in the A2 parent, implying that one of the A2 founding haplotypes was associated with mating type determination. Consistent with this result, segregation patterns for SNPs in the MTR resembled the behavior of loci in the pseudoautosomal

(conserved) regions of heteromorphic sex chromosomes (e.g. XY or ZW; (Clark, 1988)), where the A2 parent corresponded to the heterogametic, male sex. These results suggest that in populations of *P. capsici* with few founders, heterozygosity in the MTR will be maintained despite inbreeding, proportional to LD between the mating type factor(s) and the rest of the genome.

These findings, which represent the first genomic analysis of mating type in a *Phytophthora* species, are consistent with the existing models of heterozygosity versus homozygosity at a single locus as determinant of mating type (Sansome, 1980; Fabritius & Judelson, 1997). However, our analysis does not demonstrate that heterozygosity *per se* confers the A2 mating type, nor does our analysis preclude the presence of heteromorphic mating type chromosomes in *P. capsici*. We applied stringent SNP filters to obtain a high quality set of markers, likely discarding SNPs located in regions of structural variation (i.e. duplications, deletions, repeats). Indeed, early cytological work supports heterozygosity for a reciprocal translocation in association with mating type in *P. capsici* and numerous *Phytophthora* species, posited as a mechanism to suppress local recombination (Sansome, 1976). Given that chromosomal heteromorphism has arisen in diverse taxa as a consequence of suppressed recombination between sex-determining chromosomes (Charlesworth, 2013; Bachtrog, 2013), future studies will investigate the association of structural variation and recombination suppression with mating type determination in *P. capsici*.

Materials and Methods

Isolate and DNA collection. In 2008, a restricted access research field at Cornell University's New York Agricultural Experiment Station in Geneva NY, with no prior history of

Phytophthora blight, was inoculated with two NY isolates of *P. capsici*, 0664-1 (A1) and 06180-4 (A2), of opposite mating types, as described in Dunn et al (2014). From 2009-13, the field was planted with susceptible crop species, and each year the pathogen was isolated from infected plant material, cultured on PARPH medium (Dunn et al., 2014). Once in pure culture, a single zoospore isolate was obtained (Dunn et al., 2014), and species identity was confirmed with PCR using species specific primers as previously described (Zhang *et al.*, 2008; Dunn *et al.*, 2010).

Isolates collected in 2009-12 were obtained from storage; isolates from 2013 were unique to this study and were collected from infected pumpkin plants (variety Howden Biggie). Single oospore progeny (n=46) from an *in vitro* cross between the founding parents were obtained from storage (Dunn et al., 2014). To revive isolates from storage, several plugs from each storage tube were plated on PARPH media. After less than one week, actively growing cultures were transferred to new PARP or PARPH medium.

Mycelia were harvested for DNA extraction as previously described (Dunn et al., 2010), except that sterile 10% clarified V8 (CV8) broth (Skidmore *et al.*, 1984) was used instead of sterile potato dextrose broth. For each isolate, mycelia were grown in Petri plates containing CV8 broth for less than 1 week, vacuum filtered, and 90-110 mg of tissue were placed in 2ml centrifuge tubes and stored at -80C until DNA extraction. DNA was extracted using the DNeasy Plant Mini kit (Qiagen, Valencia, CA) according to manufacturer's instructions except that mycelial tissue was ground using sterile ball bearings and a TissueLyser (Qiagen, Valencia, CA) as previously described (Dunn et al., 2010).

Mating type was determined as previously described (Dunn et al., 2010). Briefly, each isolate was grown on separate unclarified V8 agar with known A1 and A2 isolates, respectively. After at least one week of growth, the plates were assessed microscopically for the presence of

oospores. For each trial, the A1 and A2 tester isolates were grown in isolation and on the same plate as negative and positive controls, respectively. We obtained mating type designations for isolates from years 2009-12 and the *in vitro* F_1 from Dunn et al. (2014).

Genotyping. All DNA samples were submitted to the Institute of Genomic Diversity at Cornell University for 96-plex GBS as previously described (Elshire et al., 2011). In brief, each sample was digested with *Ape*KI, followed by adapter ligation, and samples were pooled prior to 100bp single-end sequencing with Illumina HiSeq 2000/2500 (Elshire et al., 2011). To validate experimental procedures, DNA samples from the parental isolates were included with each sequencing plate (except in one instance). The parental isolates were sequenced initially at a higher sequencing depth (12-plex).

Genotypes were called for all isolates simultaneously using the TASSEL 3.0.173 pipeline (Glaubitz et al., 2014). This process involves aligning unique reads, trimmed to 64bp, to the reference genome (Lamour et al., 2012) and mitochondrial (courtesy of Martin, F., USDA-ARS) assemblies, and associating sequence reads with the corresponding individual by barcode identification to call SNPs (Glaubitz et al., 2014). The Burrows-Wheeler alignment (v.0.7.8) algorithm bwa-aln with default parameters (Li & Durbin, 2009) was used to align sequence tags to the reference genome (Lamour et al., 2012). To reduce downstream SNP artifacts due to poor sequencing alignment, reads with a mapping quality <30 were removed. Default parameters were otherwise used in TASSEL, with two exceptions: 1) Only sequence tags present >10 times were used to call SNPs; and 2) SNPs were output in variant call format (VCF), with up to 4 alleles retained per locus, using the tbt2vcfplugin. Genotypes were assigned and genotype likelihoods were calculated as described in (Hyma et al., 2015).

Individual and SNP Quality Control. Individuals with more than 40% missing data were removed from analysis. To mitigate heterozygote undercalling due to low sequence coverage, genotypes with depths <5 reads were set to missing using a custom python script. Subsequently, we utilized VCFtools version 1.14 (Danecek et al., 2011) to retain SNPs which met the following criteria: 1) Genomic; 2) <20% missing data; 3) Mean read depth ≥10; 4) Mean read depth <50; 5) Bi-allelic; and 6) Minor allele frequency (MAF)≥0.05. Additionally, indels were removed.

To remove isolates with likely ploidy variation, we assessed allele depth ratios for each isolate, where the allele depth ratio was defined as the ratio of the major allele to the total allele depth at a heterozygous locus (Yoshida *et al.*, 2013; Rosenblum *et al.*, 2013; Li *et al.*, 2015). Allele depths were extracted from the VCF file using a custom python script to analyze the distribution of allele depth ratios for each individual across all SNPs.

Post clone-correction (see below) and allele depth outlier removal, SNPs were further filtered in R version 3.2.3 (R Core Team, 2015). SNPs with heterozygosity rates >90% among all isolates (clone-corrected and parental replicates) were removed and/or average allele depth ratios <0.2 or >0.8. Only SNPs within scaffolds containing more than 300 kb of sequence, covering ~48MB (~75% of the sequenced genome), were retained. We defined the minor allele as the least frequent allele in the clone-corrected field population.

Multiple sequencing runs of the parental isolates were used to define consensus genotypes for each parent using the majority rule: sites where ≥50% of calls were missing or where disparate genotype calls were equally frequent were set to missing.

Identifying a clone-correction threshold. To establish a maximum similarity threshold to define unique genotypes, the genetic similarity of all sequencing runs of the parental isolates

were compared. Similarity was defined as identity-by-state (IBS), the proportion of alleles shared between two isolates at non-missing SNPs. Parental replicates represented both biological (different mycelial harvests and/or independent cultures) and technical replicates (same DNA sample), thereby capturing variation associated with culture transfers, mycelial harvests, DNA extractions and sequencing runs (Table B.S1). Based on the variation between parental replicates, individuals more than 95% similar to each other were considered clones, and one randomly selected individual from each clonal group was retained in the clone-corrected data set.

All analyses, if not otherwise specified, were performed in R using custom scripts.

Population Structure. Principal component analysis was performed on a scaled and centered genotype matrix in the R package pcaMethods (Stacklies *et al.*, 2007), using the *nipalsPCA* method to account for the small amount of missing data (method='nipals', center=TRUE, scale='uv'). This method was used for all PCAs performed. To estimate pairwise differentiation between years, we used Weir and Cockerham's (1984) F_{ST} measure (Weir & Cockerham, 1984), which weights allele frequency and variance estimates by population size, implemented in the R package *StAMMP* with the *stamppFST* function (Pembleton *et al.*, 2013). We performed 1000 permutations of the SNP set to assess if F_{ST} estimates were significantly greater than zero (Weir & Cockerham, 1984; Pembleton *et al.*, 2013).

Measures of inbreeding. We used the canonical method-of-moments estimator of the individual inbreeding coefficient, $F=1-H_o/H_e$, where H_o is the observed individual heterozygosity, and H_e is the expected heterozygosity given allele frequencies in a reference population assumed to be at HWE (Purcell *et al.*, 2007; Keller *et al.*, 2011). We utilized allele frequencies in the *in vitro* F_1 to define expected heterozygosity, a theoretical equivalent to evaluating F with respect to the parental allele frequencies (Wang, 2014). For each isolate, F was

calculated with respect to non-missing genotypes only. To compare average F between years, a pairwise t-test was implemented in R with *pairwise.t.test* (pool.sd=FALSE, paired=FALSE, p.adjust.method='bonferroni').

Heterozygosity was defined as the number of isolates with a heterozygous genotype at each SNP divided by the total number of non-missing genotype calls. Minor allele frequency (MAF) was defined as the number of minor alleles present at each SNP divided by the total number of non-missing chromosomes (number of non-missing genotype calls multiplied by two). Heterozygosity and MAF distributions for each year and the *in vitro* F_1 were graphically assessed using the *density* function in R, where the minor allele was defined as the allele with the lowest frequency in the field population.

For each individual, we calculated the proportion of MEs, defined as the ratio of MEs to the total number of non-missing tested sites, analogous to the PLINK implementation (–mendel) (Purcell et al., 2007). An ME was defined as a genotype inconsistent with the individual being an F_1 derived from the two founding parental isolates (Purcell et al., 2007). An isolate with a proportion MEs exceeding the *in vitro* F_1 mean by 3*s* was classified as field inbred and otherwise as field F_1.

Genome scan for allele frequency differentiation. To detect regions of differentiation between the *in vitro* F_1, field F_1, and inbred isolates, we performed a Fisher's exact test of allele counts for all pairwise comparisons, using the *fisher.test* function in R. *P*-values were adjusted for multiple testing using the Benjamini and Hochberg (1995) procedure, implemented with the *p.adjust* function, at a false discovery rate (FDR) of 10% (Wright, 1992; Benjamini & Hochberg, 1995). Significant SNPs were retained in further analyses only if another SNP within 200 kb also surpassed the significance threshold.

We compared the F_{ST} distribution of significantly differentiated SNPs within ROIs to the genome-wide F_{ST} distribution according to Lewontin and Krakauer (1973). Here, F_{ST} was defined as, $F_{ST} = \frac{(\Delta p)^2}{p_0(1-p_0)}$, where p_0 is the frequency of the minor allele in the field F_1, and Δp is the difference in allele frequency between the field F_1 and field inbred subpopulations.

Haploview (Barrett *et al.*, 2005) was used to estimate pairwise LD (r^2) between SNPs in scaffolds containing ROIs.

Haplotyping. As the population was established by two parental strains, assuming no mutation, all isolates were by definition combinations of the founding parental haplotypes. Therefore, we took a deterministic approach to phasing, akin to utilizing trio information to phase parental genotypes (Browning & Browning, 2011). Haplotyping in regions of interest was further facilitated by the fact that either or both parents were homozygous, with the homozygous genotype assumed to represent a founding parental haplotype. We showed that this was a valid assumption by analyzing early replicates of the parental genotypes that represented the "ancestral" heterozygous genotype in a specific region (Figure C.S13) and by comparison to homozygous genotypes of selfed isolates (data not shown). We used the homozygous parental stretches (haplotypes) to deduce the other haplotypes from consensus genotypes for the expected genotypic classes. Progeny membership in a genotypic class was defined by *k*-means clustering using the *kmeans* function in R (centers=8, n.iter=1000, nstart=100). To further refine clusters and remove recombinant isolates, we calculated local pairwise relatedness, defined as IBS, between isolates within a cluster, and removed isolates that shared on average less than 90% IBS with the respective cluster members. Next, we defined the consensus genotype based on the refined clusters utilizing the majority rule (see "Identifying a clone-correction threshold"), and heterozygous genotypes within haplotypes were set to missing.

To determine the haplotype composition of each isolate, the three identified haplotypes in a region of interest were used to construct reference genotypes for all possible haplotype combinations (e.g. H1/H2, H1/H1). Then, the genotypic discordance (i.e. the number of mismatched genotypes) between each isolate genotype and reference genotype were calculated. The most similar reference genotype was assumed to be the correct isolate genotype if discordance was less than 25%. Otherwise, the isolate genotype was deemed "Unknown."

To create phase diagrams, haplotype tagging SNPs (SNPs which unambiguously distinguished a specific haplotype) were identified at SNPs where all haplotypes had no missing data. Individual genotypes were then classified for homozygosity or heterozygosity at each haplotype tagging SNP.

Identifying mating type associated SNPs. We performed a Fisher's exact test of allele frequency differences between isolates of opposite mating types in the field F_1. Multiple test correction was performed as above (see 'Genome scan for allele frequency differentiation').

Heterozygosity in the MTR. To test differences between the heterozygote frequency distribution in the mating type region relative to the rest of the genome, we compared the heterozygosity of genome-wide SNPs sampled in equal proportions of marker types (e.g. *AA* x *Aa*) to the mating type region using the *sample* function in R without replacement (replace=FALSE), excluding SNPs not polymorphic or with missing data in the parental isolates. The identified ME-enriched SNPs were excluded (Appendix A.1). To account for an unequal ratio of A1 to A2 mating type isolates, in each test, the A2s were down-sampled (without replacement) to equate with the A1 sample size in the respective subpopulation. We used the *wilcox.test* function in R to perform a one-sided, unpaired Wilcoxon rank sum test (alternative='less', paired=FALSE), repeated for 100 random SNP samples. Additionally,

heterozygosity distributions of the A1 and A2 isolates in each subpopulation were compared to the respective genome-wide distribution, with SNP but not isolate down-sampling.

Heterozygote excess was tested at each locus in the A1 and A2 isolates for each subpopulation, using the function *HWExact* in the R package HardyWeinberg (Wigginton *et al.*, 2005; Graffelman, 2015). This amounts to a one-sided test of HWE where heterozygote excess is the only evidence of deviation from HWE. We controlled for multiple testing as above.

Allele frequency changes in the MTR. In the MTR, the frequency of the parental tagging allele (p_a) at SNPs heterozygous in one parent and homozygous in the other, was calculated for A1 and A2 isolates separately in the parental generation, the field F_1 and the field inbreds, excluding missing genotypes. The A2 tagging SNPs were separated into two categories based on p_a with respect to mating type in the field F_1. The first case consisted of SNPs with $p_a \geq 0.3$ in the A2 isolates and $p_a \leq 0.3$ in the A1 isolates, and the second case consisted of the remaining SNPs. Expectations for p_a in theoretical F_1 and F_2 populations, for the three cases where the *a* alleles is in the haplotype background of the: 1) Y in the male sex; 2) X in the male sex; and 3) X in the female sex, were derived based on the formulas in (Clark, 1988; Allendorf *et al.*, 1994).

REFERENCES

Abdellaoui A, Hottenga JJ, de Knijff P *et al.*, 2013. Population structure, migration, and diversifying selection in the Netherlands. *European Journal of Human Genetics* **21**, 1277–1285.

Allendorf FW, Gellman WA, Thorgaard GH, 1994. Sex-linkage of two enzyme loci in Oncorhyncus mykiss (rainbow trout). *Heredity* **72 (Pt 5)**, 498–507.

Babadoost M, Pavon C, 2013. Survival of Oospores of *Phytophthora capsici* in Soil. *Plant Disease* **97**, 1478–1483.

Bachtrog D, 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature Publishing Group* **14**, 113–124.

Balloux F, 2004. Heterozygote excess in small populations and the heterozygote-excess effective population size. *Evolution* **58**, 1891.

Balloux F, Amos W, Coulson T, 2004. Does heterozygosity estimate inbreeding in real populations? *Molecular Ecology* **13**, 3021–3031.

Balloux F, Lehmann L, de Meeûs T, 2003. The population genetics of clonal and partially clonal diploids. **164**, 1635–1644.

Barrett JC, Fry B, Maller J, Daly MJ, 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics (Oxford, England)* **21**, 263–265.

Benjamini Y, Hochberg Y, 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)* **57**, 289–300.

Bowers JH, 1990. Effect of Soil Temperature and Soil-Water Matric Potential on the Survival of *Phytophthora capsici* in Natural Soil. *Plant Disease* **74**, 771.

Browning SR, Browning BL, 2011. Haplotype phasing: existing methods and new developments. *Nature Publishing Group* **12**, 703–714.

Chamnanpunt J, Shan WX, Tyler BM, 2001. High frequency mitotic gene conversion in genetic hybrids of the oomycete Phytophthora sojae. *Proceedings of the National Academy of Sciences* **98**, 14530–14535.

Charlesworth B, 2009. Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics* **10**, 195–205.

Charlesworth B, Charlesworth D, 1987. Inbreeding Depression And Its Evolutionary Consequences. *Annual Review of Ecology and Systematics* **18**, 237–268.

Charlesworth D, 2003. Effects of inbreeding on the genetic diversity of populations.

Philosophical Transactions of the Royal Society B: Biological Sciences **358**, 1051–1070.

Charlesworth D, 2013. Plant sex chromosome evolution. *Journal of Experimental Botany* **64**, 405–420.

Clark AG, 1988. The evolution of the Y chromosome with X-Y recombination. **119**, 711–720.

Danecek P, Auton A, Abecasis G *et al.*, 2011. The variant call format and VCFtools. *Bioinformatics (Oxford, England)* **27**, 2156–2158.

Dunn AR, Bruening SR, Grünwald NJ, Smart CD, 2014. Evolution of an Experimental Population of *Phytophthora capsici* in the Field. *Phytopathology* **104**, 1107–1117.

Dunn AR, Milgroom MG, Meitz JC *et al.*, 2010. Population Structure and Resistance to Mefenoxam of *Phytophthora capsici* in New York State. *Plant Disease* **94**, 1461–1468.

Elshire RJ, Glaubitz JC, Sun Q *et al.*, 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one* **6**, e19379.

Erwin DC, Ribeiro OK, 1996. *Phytophthora Diseases Worldwide*. St Paul, MN: The American Phytopathological Society.

Fabritius AL, Judelson HS, 1997. Mating-type loci segregate aberrantly in Phytophthora infestans but normally in Phytophthora parasitica: implications for models of mating-type determination. *Current genetics* **32**, 60–65.

Falconer DS, Mackay TFC, 1996. *Introduction to Quantitative Genetics*. Burnt Mill, Harlow, Essex, England: Longman.

Felsenstein J, 1971. Inbreeding and variance effective numbers in populations with overlapping generations. *Genetics* **68**, 581–597.

Forche A, Abbey D, Pisithkul T *et al.*, 2011. Stress Alters Rates and Types of Loss of Heterozygosity in Candida albicans. *mBio* **2**, e00129–11–e00129–11.

Galtier N, Depaulis F, BARTON NH, 2000. Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. **155**, 981–987.

Glaubitz JC, Casstevens TM, Lu F *et al.*, 2014. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PloS one* **9**, e90346.

Graffelman J, 2015. Exploring Diallelic Genetic Markers: The Hardy Weinberg Package. *Journal of statistical software* **64**.

Granke LL, Quesada-Ocampo L, Lamour K, 2012. Advances in research on *Phytophthora capsici* on vegetable crops in the United States. *Plant Disease* **96**, 1588–1600.

Granke LL, Windstam ST, Hoch HC, Smart CD, Hausbeck MK, 2009. Dispersal and movement mechanisms of Phytophthora capsici sporangia. *Phytopathology* **99**, 1258–1264.

Grünwald NJ, Garbelotto M, Goss EM, Heungens K, Prospero S, 2012. Emergence of the sudden oak death pathogen Phytophthora ramorum. *Trends in Microbiology* **20**, 131–138.

Grünwald NJ, Goodwin SB, Milgroom MG, Fry WE, 2003. Analysis of genotypic diversity data for populations of microorganisms. *Phytopathology* **93**, 738–746.

Hairston NG Jr, De Stasio BT Jr, 1988. Rate of evolution slowed by a dormant propagule pool. *Nature* **336**, 239–242.

Hartl DL, Clark AG, 2007. *Principles of Population Genetics*. Sunderland, MA: Sinauer Associates, Inc. Publishers.

Hausbeck MK, Lamour KH, 2004. *Phytophthora capsici* on vegetable crops: research progress and management challenges. *Plant Disease* **88**, 1292–1303.

Hurtado-Gonzales OP, Lamour KH, 2009. Evidence for inbreeding and apomixis in close crosses of Phytophthora capsici. *Plant Pathology* **58**, 715–722.

Hyma KE, Barba P, Wang M *et al.*, 2015. Heterozygous Mapping Strategy (HetMappS) for High Resolution Genotyping-By-Sequencing Markers: A Case Study in Grapevine (NA Tinker, Ed,). *PloS one* **10**, e0134880–31.

Jorde PE, Ryman N, 1995. Temporal allele frequency change and estimation of effective size in populations with overlapping generations. **139**, 1077–1090.

Kardos M, Luikart G, Allendorf FW, 2015. Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. *Heredity* **115**, 63–72.

Kasuga T, Bui M, Bernhardt E *et al.*, 2016. Host-induced aneuploidy and phenotypic diversification in the Sudden Oak Death pathogen Phytophthora ramorum. *BMC Genomics*, 1–17.

Keller MC, Visscher PM, Goddard ME, 2011. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. **189**, 237–249.

Kirkpatrick M, Jarne P, 2000. The Effects of a Bottleneck on Inbreeding Depression and the Genetic Load. *The American Naturalist* **155**, 154–167.

Ko W, 1988. Hormonal heterothallism and homothallism in *Phytophthora*. *Annual Review of Phytopathology* **26**, 57–73.

Kondrashov AS, 1988. Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440.

Lamour KH, Hausbeck MK, 2000. Mefenoxam Insensitivity and the Sexual Stage of Phytophthora capsici in Michigan Cucurbit Fields. *Phytopathology* **90**, 396–400.

Lamour KH, Hausbeck MK, 2001. Investigating the Spatiotemporal Genetic Structure of

Phytophthora capsici in Michigan. *Phytopathology* **91**, 973–980.

Lamour KH, Hausbeck MK, 2003. Effect of crop rotation on the survival of *Phytophthora capsici* in Michigan. *Plant Disease* **87**, 841–845.

Lamour KH, Mudge J, Gobena D *et al.*, 2012. Genome Sequencing and Mapping Reveal Loss of Heterozygosity as a Mechanism for Rapid Adaptation in the Vegetable Pathogen *Phytophthora capsici*. *Molecular Plant-Microbe Interactions* **25**, 1350–1360.

Lewontin RC, Krakauer J, 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. **74**, 175–195.

Li H, Durbin R, 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* **25**, 1754–1760.

Li Y, Zhou Q, Qian K, van der Lee T, Huang S, 2015. Successful asexual lineages of the Irish potato Famine pathogen are triploid. *bioRxiv*, 024596.

Luikart G, Cornuet JM, 1999. Estimating the effective number of breeders from heterozygote excess in progeny. **151**, 1211–1216.

Magwene PM, Kayıkçı Ö, Granek JA, Reininga JM, Scholl Z, Murray D, 2011. Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 1987–1992.

Mandegar MA, Otto SP, 2007. Mitotic recombination counteracts the benefits of genetic segregation. *Proceedings of the Royal Society of London B: Biological Sciences* **274**, 1301–1307.

Marshall AR, Knudsen KL, Allendorf FW, 2004. Linkage disequilibrium between the pseudoautosomal PEPB-1 locus and the sex-determining region of chinook salmon. *Heredity* **93**, 85–97.

Meirmans PG, van Tienderen PH, 2004. genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**, 792–794.

Milgroom MG, 1996. Recombination and the multilocus structure of fungal populations. *Annual Review of Phytopathology* **34**, 457–477.

Nunney L, 2002. The Effective Size of Annual Plant Populations: The Interaction of a Seed Bank with Fluctuating Population Size in Maintaining Genetic Variation. *The American Naturalist* **160**, 195–204.

Pembleton LW, Cogan NOI, Forster JW, 2013. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources* **13**, 946–952.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D, 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**, 904–909.

Pudovkin AI, Zaykin DV, Hedgecock D, 1996. On the potential for estimating the effective number of breeders from heterozygote-excess in progeny. **144**, 383–387.

Purcell S, Neale B, Todd-Brown K *et al.*, 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **81**, 559–575.

R Core Team, 2015. R: A language and environment for statistical computing.

Robertson A, 1965. The interpretation of genotypic ratios in domestic animal populations. *Animal Production* **7**, 319–324.

Rogstad SH, Keane B, Beresh J, 2002. Genetic variation across VNTR loci in central North American Taraxacum surveyed at different spatial scales. *Plant Ecology* **161**, 111–121.

Rosenblum EB, James TY, Zamudio KR *et al.*, 2013. Complex history of the amphibian-killing chytrid fungus revealed with genome resequencing data. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 9385–9390.

Sansome E, 1976. Gametangial meiosis in Phytophthora capsici. *Canadian Journal of Botany* **54**, 1535–1545.

Sansome E, 1980. Reciprocal translocation heterozygosity in heterothallic species of Phytophthora and its significance. *Transactions of the British Mycological Society* **74**, 175–185.

Satour MM, Butler EE, 1967. A root and crown rot of tomato caused by *Phytophthora capsici* and *P. parasitica*. *Phytopathology*, 510–515.

Satour MM, Butler EE, 1968. *Comparative morphological and physiological studies of progenies from intraspecific matings of Phytophthora capsici*. Phytopathology.

Shattock RC, 1986. Genetics of *Phytophthora infestans*: Characterization of Single-Oospore Cultures from A1 Isolates Induced to Self by Intraspecific Stimulation. *Phytopathology* **76**, 407.

Skidmore DI, Shattock RC, Shaw DS, 1984. Oospores in cultures of *Phytophthora infestans* resulting from selfing induced by the presence of *P. drechsleri* isolated from blighted potato foliage. *Plant Pathology* **33**, 173–183.

Stacklies W, Redestig H, Scholz M, Walther D, Selbig J, 2007. pcaMethods--a bioconductor package providing PCA methods for incomplete data. *Bioinformatics (Oxford, England)* **23**, 1164–1167.

Templeton AR, Levin DA, 1979. Evolutionary Consequences of Seed Pools. *The American Naturalist* **114**, 232–249.

Tian D, Babadoost M, 2004. Host range of *Phytophthora capsici* from pumpkin and pathogenicity of isolates. *Plant Disease* **88**, 485–489.

Uchida JY, Aragaki M, 1980. Chemical stimulation of oospore formation in *Phytophthora capsici*. *Mycologia* **72**, 1103.

Wang J, 2014. Marker-based estimates of relatedness and inbreeding coefficients: an assessment of current methods. *Journal of Evolutionary Biology* **27**, 518–530.

Waples RS, 2006. Seed Banks, Salmon, and Sleeping Genes: Effective Population Size in Semelparous, Age-Structured Species with Fluctuating Abundance. *The American Naturalist* **167**, 118–135.

Waples RS, 2014. Testing for Hardy-Weinberg Proportions: Have We Lost the Plot? *Journal of Heredity* **106**, 1–19.

Weir BS, Cockerham CC, 1984. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**, 1358.

Wigginton JE, Cutler DJ, Abecasis GR, 2005. A Note on Exact Tests of Hardy-Weinberg Equilibrium. *The American Journal of Human Genetics* **76**, 887–893.

Wright S, 1921. Systems of Mating. II. The Effects of Inbreeding on the Genetic Composition of a Population. **6**, 124–143.

Wright SP, 1992. Adjusted p-values for simultaneous inference. *Biometrics* **48**, 1005.

Yoshida K, Schuenemann VJ, Cano LM *et al.*, 2013. The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *eLife* **2**, 403–25.

Zhang N, McCarthy ML, Smart CD, 2008. A macroarray system for the detection of fungal and oomycete pathogens of solanaceous crops. *Plant Disease* **92**, 953–960.

CONCLUSION

In the present study, I utilized genome-wide SNP genotyping to characterize the genetic structure of a closed, biparental field population of *P. capsici* over time (Dunn *et al.*, 2014). This unique experimental design inherently imposed selective pressures on the pathogen population, in that isolates were required to both withstand variable environmental conditions and successfully infect a plant. At the same time, initiation of the field population with only two parental isolates, implied that generations beyond the F_1 were inbred, and thus more homozygous. Indeed, by analyzing changes in individual heterozygosity over time, I showed that inbred isolates were recovered in the third year after establishment (2011). Furthermore, I demonstrated that oospore survival in the soil mitigated heterozygosity declines due to inbreeding. Acting akin to the seed bank of an annual plant species, in natural infestations, multi-year oospore survival in the soil is likely to serve as a genetic reservoir, in addition to ensuring the persistence of pathogen populations in the absence of susceptible hosts (Lamour & Hausbeck, 2003; Dunn *et al.*, 2014).

While not explicitly addressed in the present study, if increases in homozygosity corresponded to fitness declines (in terms of pathogenicity and/or fecundity), long-term survival of outbred oospores formed during the initial infestation, could play a large role in maintaining aggressive, pathogenic isolates in a natural population over time. In support of this hypothesis, a very low rate of self-fertilization in both *in vitro* and field populations, despite hermaphroditism, implies a cost to homozygosity, as observed in essentially all outcrossing species (Charlesworth & Charlesworth, 1987; Falconer & Mackay, 1996). On the other hand, evidence for mitotic LOH in *Phytophthora* species in the present study and numerous prior studies (Chamnanpunt *et al.*,

2001; Grünwald *et al.*, 2012; Lamour *et al.*, 2012; Kasuga *et al.*, 2016), suggests that *Phytophthora* genomes not only tolerate declines in heterozygosity, but that this process may aid adaptation. Reconciling potential fitness costs with incidence of mitotic LOH will provide key insights into the role of asexual reproduction in pathogen adaptation.

Given the inferred cost of homozygosity, the observed large shifts in haplotype frequency in specific regions of the genome, may be a result of natural selection acting upon recessive deleterious alleles in the homozygous state. Simultaneously, numerous processes, such as reductions in population size, asexual reproduction, and inbreeding, likely decreased the effective population size relative to the census population size, consequently increasing the magnitude of genetic drift. In particular, in contrast to natural populations of *P. capsici* (Lamour & Hausbeck, 2001; Dunn *et al.*, 2010), we consistently observed skewed mating type ratios in favor of the A2 mating type in the biparental field population as well as in the *in vitro* F_1. As each sex, here mating type, contributes 50% of the total alleles to the next generation, skewed sex ratios likely reduced the effective population size (Hartl & Clark, 2007). However, additional work is required to address the interaction between skewed mating type ratios, asexual reproduction, and stochastic processes (e.g. proximity to a mate), on the contribution of a given isolate to the so-called oospore bank.

Revelation of a mating type determination system in which one of the A2 founding haplotypes is consistently associated with the A2 mating type, may have implications for observation of both a low rate of self-fertilization and biased mating type ratios. Furthermore, we show that the mating system resulted in allele and genotypic frequency signatures distinct from the rest of the genome among inbred isolates. Future work will elucidate the molecular basis of

mating type determination and whether or not structural variation is present at the mating type locus.

Final thoughts

Much of molecular plant pathology has focused on the interaction between a limited number of pathogen strains and host cultivars. Experiments are often performed in highly controlled conditions, not representative of a field environment (Cai *et al.*, 2012). While these controlled experiments provide the basis for our understanding of plant-microbe interactions, they neglect the role of pathogen populations and environment in the coevolution of plant pathogens with their respective plant hosts. Yet, is has long been acknowledged that both environment and population biology, defined by processes including gene flow, asexual reproductive rate, mating system, and mutation rate, have significant implications for the durability of host resistance and disease management (Goodwin, 1997; Mcdonald & Linde, 2002). At the same time, analysis of pathogen genomes has revealed genome architecture conducive to rapid adaption (Raffaele & Kamoun, 2012; Seidl & Thomma, 2014), and the presence of genetic variation within clonal lineages. Progressively declining sequencing costs and advancements in phenotyping technologies provide an unprecedented opportunity to integrate population genomic analysis with fine-scale investigation of plant-microbe interactions (Grünwald *et al.*, 2015).

REFERENCES

Cai G, Restrepo S, Myers K *et al.*, 2012. Gene profiling in partially resistant and susceptible near-isogenic tomatoes in response to late blight in the field. *Molecular Plant Pathology* **14**, 171–184.

Chamnanpunt J, Shan WX, Tyler BM, 2001. High frequency mitotic gene conversion in genetic hybrids of the oomycete Phytophthora sojae. *Proceedings of the National Academy of Sciences* **98**, 14530–14535.

Charlesworth B, Charlesworth D, 1987. Inbreeding Depression And Its Evolutionary Consequences. *Annual Review of Ecology and Systematics* **18**, 237–268.

Dunn AR, Bruening SR, Grünwald NJ, Smart CD, 2014. Evolution of an Experimental Population of *Phytophthora capsici* in the Field. *Phytopathology* **104**, 1107–1117.

Dunn AR, Milgroom MG, Meitz JC *et al.*, 2010. Population Structure and Resistance to Mefenoxam of *Phytophthora capsici* in New York State. *Plant Disease* **94**, 1461–1468.

Falconer DS, Mackay TFC, 1996. *Introduction to Quantitative Genetics*. Burnt Mill, Harlow, Essex, England: Longman.

Goodwin SB, 1997. The Population Genetics of *Phytophthora*. *Phytopathology* **87**, 462–473.

Grünwald NJ, Garbelotto M, Goss EM, Heungens K, Prospero S, 2012. Emergence of the sudden oak death pathogen Phytophthora ramorum. *Trends in Microbiology* **20**, 131–138.

Grünwald NJ, Mcdonald BA, Milgroom MG, 2015. Population Genomics of Fungal and Oomycete Pathogens. *Annual Review of Phytopathology* **54**, annurev–phyto–080614–115913–24.

Hartl DL, Clark AG, 2007. *Principles of Population Genetics*. Sunderland, MA: Sinauer Associates, Inc. Publishers.

Kasuga T, Bui M, Bernhardt E *et al.*, 2016. Host-induced aneuploidy and phenotypic diversification in the Sudden Oak Death pathogen Phytophthora ramorum. *BMC Genomics*, 1–17.

Lamour KH, Hausbeck MK, 2001. Investigating the Spatiotemporal Genetic Structure of *Phytophthora capsici* in Michigan. *Phytopathology* **91**, 973–980.

Lamour KH, Hausbeck MK, 2003. Effect of crop rotation on the survival of *Phytophthora capsici* in Michigan. *Plant Disease* **87**, 841–845.

Lamour KH, Mudge J, Gobena D *et al.*, 2012. Genome Sequencing and Mapping Reveal Loss of Heterozygosity as a Mechanism for Rapid Adaptation in the Vegetable Pathogen *Phytophthora capsici*. *Molecular Plant-Microbe Interactions* **25**, 1350–1360.

Mcdonald BA, Linde C, 2002. Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology* **40**, 349–379.

Raffaele S, Kamoun S, 2012. Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Publishing Group* **10**, 417–430.

Seidl MF, Thomma BPHJ, 2014. Sex or no sex: Evolutionary adaptation occurs regardless. *BioEssays* **36**, 335–345.

APPENDIX A

SUPPLEMENTARY TEXT

A.1. Correcting for mitotic LOH in ME estimates

Initially, the mean proportion MEs for the *in vitro* F_1 (5.77%) and empirically defined field F_1 isolates (6.34%) exceeded our estimated genotyping error rate of 3% (by 2.77% and 3.34%, respectively). Motivated by this observation, we assessed whether specific SNPs were contributing disproportionately to overall ME estimates, by calculating the proportion MEs for each SNP in both the combined *in vitro* F_1 and field F_1 subpopulations (n=143). The maximum number of tested sites for each individual consisted of the consensus parental genotypes excluding double heterozygous SNPs data, which are uninformative in assessing MEs, and SNPs with missing parental data. An isolate with a proportion MEs exceeding the *in vitro* F_1 mean by three standard deviations (>10.65%) was classified as field inbred and otherwise as field F_1.

As a result, we identified 848 ME-enriched SNPs, defined as a SNP with greater than 10% MEs (equivalent to 15 isolates with a ME). While some ME-enriched SNPs were isolated and randomly distributed throughout the genome, ME-enriched SNPs appeared in clusters in several instances (Figure C.S6), suggestive of underlying biological factors rather than sequencing or genotyping error. We show below that these events occurred post-field inoculation but prior to genotyping of the parents, resulting in homozygous parental genotypes discordant with segregation of four haplotypes in the field population. Additionally, ME-enriched regions were associated with differentiation between the *in vitro* F_1 and field progeny in six of the seven cases.

A.2. Evidence for LOH events: segregation in the in vitro F_1 and field population provide evidence for loss of heterozygosity (LOH) events

The *in vitro* F_1 were representative of the field F_1 in terms of inbreeding coefficient, site heterozygosity (data not shown) and MAF distributions (Figure C.S5 B). To further evaluate the extent to which the *in vitro* F_1 equated to the field F_1, we performed a Fisher's exact test of allele frequency differences (Weir, 1996) between these two subpopulations at each SNP. We utilized a Fisher's exact rather than chi-square test due to the 0.08% of cases where the expected allele counts were <5 in all pairwise comparisons.

Genome-wide, allele frequencies between the *in vitro* F_1 and the field F_1 were similar, except for six regions (in scaffolds 8, 19, 26, 33, 35, 55) with 5% to 68% of SNPs within each region exceeding the multiple-test correction threshold (Figure C.S7 A; Table B.S3). These six regions were likewise highly differentiated between the *in vitro* F_1 and field inbred subpopulations (Figure C.S7 B). Regions of differentiation co-localized with the identified ME-enriched SNP clusters (Figure C.S8 B & E).

For each region, we first performed PCA on all isolates and the consensus parental genotypes, using only the SNPs within the minimum and maximum significantly differentiated SNPs in each scaffold. In each PCA, we observed four primary clusters, indicative of four distinct genotypes (Figure C.S9). This was in accordance with expectations for an F_1 derived from heterozygous parents, where four segregating haplotypes result in four genotypes in the progeny. However, we observed very low heterozygosity in either or both the parents in these regions, suggesting fewer than four founding haplotypes (Figure C.S8 C & F).

To understand the discrepancy between parental homozygosity and observed segregation in the field F_1, we locally phased all isolates in each region, excluding R-19, utilizing a deterministic approach (see Methods). We then counted the number of distinct haplotypes in both the parents and the field F_1. The field F_1 presented four haplotypes (H1, H2, H3 and H4) in each region, whereas the parental genotypes provided evidence for only two to three haplotypes (Table B.S3). For example, in the scaffold 26 region (R-26), if the parental isolates used to inoculate the field were as sequenced (A1 parent=H1/H2 and A2 parent=H3/H3), we would expect only H1/H3 and H2/H3 progeny. Yet, we observed H1/H4 and H2/H4 genotypes in the field F_1 and the *in vitro* F_1, resulting in an excess of MEs in R-26 for both the field F_1 and *in vitro* F_1 (Figure C.S8 A and see C.S10). The simplest explanation for this discrepancy was that the sequenced A2 parental isolate underwent a mitotic loss of heterozygosity (LOH) event in R-26 after the field inoculation and collection of *in vitro* progeny (i.e. during culture prior to sequencing). Thus, in R-26, the genotype of the sequenced isolate (H3/H3) differed from the inferred genotype (H3/H4) of the isolate used to found the field and *in vitro* subpopulations. Segregation in the scaffold 8 region (R-8) followed a similar pattern to segregation in R-26 (Figure C.S8 D-F; Table B.S4), consistent with both scaffolds being adjacent in linkage group 8 (Lamour *et al.*, 2012).

In the scaffold 19, 33, 35, and 55 regions, the *in vitro* F_1 were present in only two of the four PCA clusters (Figure C.S9 A), and lacked one haplotype (H2) relative to the field F_1 (Table B.S4; Figures C.S10-12). Akin to R-8 and R-26, A1 parent homozygosity (H1/H1) in these four regions conflicted with the four observed genotypes, supporting incidence of LOH in the A1 parental culture in these regions. As the most parsimonious explanation for the presence of H2 in the field F_1, but not the *in vitro* F_1, we hypothesized that the A1 parental LOH event occurred in culture after the field inoculation (2008), but predated collection of the *in vitro* progeny (2010). This hypothesis was supported by markedly higher proportions of MEs in the field F_1 relative to the *in vitro* F_1 in the scaffold 19, 33, 35, and 55 regions (Figure C.S8). Observation of the same pattern in these four regions, was consistent with presence of the corresponding scaffolds in linkage group 16 (Lamour *et al.*, 2012).

In R-35 both the A1 and A2 parents were homozygous (H1/H1 and H3/H3, respectively), which would result in only H1/H3 progeny. Here, since we observed haplotype H4 (A2 parent haplotype) in both the *in vitro* and the field F_1, the A2 parent LOH event would have necessarily occurred post field inoculation and *in vitro* collection. To investigate the timing of this LOH event, we used the fact that the parental replicates represented multiple distinct culture time points (from 2013 to 2014), to compare the genotypes of the parental isolates in R-35 across culture time. The earliest cultures of the A2 parent (Table B.S1) were heterozygous in R-35, recapitulating the inferred H3/H4 founding genotype (Figure C.S13). Sequence-based evidence of the "ancestral" genotype strongly supported incidence of an LOH event during culture passage in 2014, after establishment of the field population and *in vitro* collection. All A1 replicates were homozygous in R-35, consistent with the A1 parent LOH event predating *in vitro* collection (2010). We also found evidence for heterozygosity in the earliest sequenced cultures of the A2 parental isolate in R-55 and in scaffold 7 concurrent with incidence of MEs in the *in vitro* F_1 and field F_1 (data not shown). Sequence based evidence and/or discordant segregation among *in vitro* and field progeny with respect to the parental genotypes associated with ME clusters, support mitotic LOH in the parental cultures as the explanation for incidence of heightened MEs in all seven ME-enriched SNP clusters.

Apart from the incidence of parental LOH in these regions, we observed skewed segregation ratios in both the field F_1 and *in vitro* F_1 in each of the six regions (Table B.S4). In fact, distorted segregation drove allele frequency differences between these two sub-populations in R-8 and R-26, rather than the incidence of parental LOH events (as was the case in the other four regions). Further, 4 to 8 *in vitro* F_1 isolates (Figures C.S10-12) were homozygous in each region (Table B.S3). (Note that these homozygous isolates were removed from calculation of MEs presented in Figure C.S8.) These specific instances of anomalous segregation, i.e. haplotype homozygosity in an F_1, could result from mitotic LOH, as likely occurred in the parental isolates due to serial culturing, or by non-Mendelian meiotic processes. In rare cases (3 to 5 isolates), we also observed homozygosity in the field F_1 in these regions (Table B.S4; Figures C.S10-13). As the number of cultures prior to sequencing was not controlled, we could not infer the relative frequency of meiotic or mitotic LOH events in the field F_1 versus *in vitro* F_1 isolates. In addition, the smaller *in vitro* F_1 sample size (n=41) may have influenced differentiation from the field F_1 in R-8 and R-26.

Due to the low frequency (<5%) of haplotype homozygosity among field F_1 isolates in these regions, aberrant LOH processes likely minimally influenced our analysis. The instances of LOH in the A1 parent that manifested in allele frequency differences between the *in vitro* F_1 and field F_1 (Rs-19, 33, 35, and 55), reflected the genomic changes occurring in culture from the time of field inoculation to collection of the *in vitro* progeny, 2008 to 2010. These four LOH tracts spanned less than 1 Mb (approximately 2% of the genome assayed), supporting negligible influence of large scale LOH events on our genome-wide analyses.

A3. Mating type associated SNPs in the field inbred

Based on the 184 SNPs associated in the F_1, the PCA of all isolates (*in vitro* and field; n=203) showed incomplete differentiation according to mating type (Figure C.16). When the Fisher's exact test was repeated in the field inbred subpopulation (n_{A1}=21 and n_{A2}=32), only SNPs within scaffolds 4 and 27 (n=53 and 20, respectively) were significantly differentiated (Figure C.S15 and Table B.S7). Utilizing only the intersection of significant SNPs, from both field F_1 and field inbred tests (n=51), the PCA more discretely separated isolates by mating type, revealing two primary clusters (Figure C.16). This result suggested that only a subset of the F_1-SNPs were tagging mating type association, but also may have been influenced by reduced power due to small sample size.

REFERENCES

Lamour KH, Mudge J, Gobena D *et al*., 2012. Genome Sequencing and Mapping Reveal Loss of Heterozygosity as a Mechanism for Rapid Adaptation in the Vegetable Pathogen *Phytophthora capsici*. *Molecular Plant-Microbe Interactions* **25**, 1350–1360.

Weir BS, 1996. *Genetic Data Analysis II*. Sunderland, MA: Sinauer Associates, Inc. Publishers.

APPENDIX B

SUPPLEMENTARY TABLES

Replicate	Short ID	Culture date	Mycelial culture date	DNA extraction date	Sequence date	Sequence Plate	Mean IBS	Culture group	Rep Group
				A1 parent					
1	Pcap4777	pre-7/2013	pre-7/2013	pre-7/2013	7/31/13	C270BACXX_3	97.65%	-	-
2	664_1	pre-9/2013	pre-9/2013	9/30/13	5/27/14	C4B1JACXX_8	98.22%	-	-
3	0664_1_T25	2/17/14	pre-9/2014	pre-9/2014	9/19/14	C507CACXX_7	98.01%	-	-
4	0664_1_T26	4/7/14	pre-9/2014	pre-9/2014	9/19/14	C507CACXX_7	97.72%	-	-
5	664_T29_1a	12/8/14	3/16/15	3/23/15	6/12/15	C6H57ANXX_7	98.48%	e	e
6	664_T29_1b	12/8/14	3/16/15	3/23/15	6/12/15	C6H57ANXX_7	98.44%	e	e
7	664_T29_2a	12/8/14	3/16/15	3/23/15	6/12/15	C6H57ANXX_7	98.49%	e	f
8	664_T29_2b	12/8/14	3/16/15	3/23/15	6/12/15	C6H57ANXX_7	98.44%	e	f
9	0664_T29_1	12/8/14	3/16/15	3/23/15	8/27/15	C6P86ANXX_1	98.47%	e	-
10	0664_T29_2	12/8/14	3/16/15	3/23/15	8/27/15	C6P86ANXX_1	98.44%	e	-
11	664_T29_1a	12/8/14	3/16/15	3/23/15	8/3/15	C6RD8ANXX_1	98.35%	e	e
12	664_T29_1b	12/8/14	3/16/15	3/23/15	8/3/15	C6RD8ANXX_1	98.25%	e	e
13	664_T29_2a	12/8/14	3/16/15	3/23/15	8/3/15	C6RD8ANXX_1	98.18%	e	f
14	664_T29_2b	12/8/14	3/16/15	3/23/15	8/3/15	C6RD8ANXX_1	98.19%	e	f
				A2 parent					
1	Pcap4778	pre-7/2013	pre-7/2013	pre-7/2013	7/31/13	C270BACXX_3	97.92%	-	-
2	6180_4	pre-9/2013	pre-9/2013	9/30/13	5/27/14	C4B1JACXX_8	97.99%	-	-
3	06180_4_T21	pre-9/2013	pre-9/2013	9/30/13	9/19/14	C507CACXX_7	97.71%	-	-
4	6180_0217a	2/17/15	3/16/15	3/23/15	6/12/15	C6H57ANXX_7	98.30%	c	d
5	6180_0217b	2/17/15	3/16/15	3/23/15	6/12/15	C6H57ANXX_7	98.39%	c	d
6	6180_012815a	1/28/15	3/16/15	3/23/15	6/12/15	C6H57ANXX_7	98.39%	d	e
-	6180_012815b	1/28/15	3/16/15	3/23/15	6/12/15	C6H57ANXX_7	-	-	-
7	06180_128	1/28/15	3/16/15	3/23/15	8/27/15	C6P86ANXX_1	98.09%	d	-
8	6180_0217a	2/17/15	3/16/15	3/23/15	8/3/15	C6RD8ANXX_1	98.03%	c	d
9	6180_0217b	2/17/15	3/16/15	3/23/15	8/3/15	C6RD8ANXX_1	98.24%	c	d
10	6180_012815a	1/28/15	3/16/15	3/23/15	8/3/15	C6RD8ANXX_1	98.24%	d	e
11	6180_012815b	1/28/15	3/16/15	3/23/15	8/3/15	C6RD8ANXX_1	98.12%	d	e

Table S1. Replicates of the parental isolates.

	Sample					
Variables	2009	2010	2011	2012	2013	Total field
	Not clone-corrected					
MT						
A1	14	21	21	11	17*	84
A2	26	34	32	24	31	147
$\chi 2$ *P*-value**	**0.058**	**0.080**	0.131	**0.028**	**<0.001**	**<0.001**
Total	40	55	53	35	48	231
	Clone-corrected					
MT						
A1	13	16	18	9	9	65
A2	23	23	27	7	14	94
$\chi 2$ *P*-value**	**0.096**	0.262	0.180	0.617	0.297	**0.021**
Total	36	39	45	16	23	159
Unique genotypes (%)	90.00	70.91	84.91	45.71	47.92	68.83

*Excluding the isolate, 13PF_29A, which exhibited a skewed allele depth ratio distribution.
**Bold indicates significance ($\alpha<0.1$)

Table S2. Counts of A1 and A2 mating types among non-clone-corrected and clone-corrected isolates with respect to year.

Region	Scaffold	Minimum Significant Position (bp)	Maximum Significant Position (bp)	ROI size (bp)	Significant SNPs (#)	Significant SNPs in region (%)	Density of Significant SNPs (SNP/kb)	Linkage group
R-8	8	341,231	1,121,521	780,290	6	5.22	0.01	8
R-19	19	704,365	888,453	184,088	23	67.65	0.12	16
R-26	26	101,162	618,561	517,399	14	16.28	0.03	8
R-33	33	54,561	181,690	127,129	17	45.95	0.13	16
R-35	35	142,871	508,327	365,456	22	31.88	0.06	16
R-55	55	85,182	383,472	298,290	29	59.18	0.1	16

Table S3. Regions of differentiation between field F_1 and *in vitro* F_1 isolates associated with incidence of Mendelian errors.

			Genotype counts*									Haplotype counts***			
	H1/H2	H3/H4	H1/H3	H1/H4	H2/H3	H2/H4	*H1/H1*	*H2/H2*	*H3/H3*	*H4/H4*	*Unknown***	**H1**	**H2**	**H3**	**H4**
R-8 (sc8)															
Field F1			27	25	18	22	*1*	*1*	*1*		9	54	42	47	47
Observed Parents	**A1**								***A2***			1	1	1	0
in vitro F1			12	6	15	4	*1*		*3*			20	19	33	10
R-26 (sc26)															
Field F1			15	24	27	32		*2*	*1*		3	39	63	44	56
Observed Parents	**A1**								***A2***			1	1	1	0
in vitro F1			13	5	14	5	*1*		*3*			20	19	33	10
R-33 (sc33)															
Field F1	*1*		27	28	20	18	*1*	*1*	*2*		6	58	41	51	46
Observed Parents		**A2**					***A1***					1	1	1	0
in vitro F1			13	23	0	1	*3*			*1*		42	1	13	26
R-35 (sc35)															
Field F1	*1*		17	33	21	24	*1*	*1*	*2*	*1*	3	53	48	42	59
Observed Parents							***A1***		***A2***			1	1	1	0
in vitro F1			19	12	0	0	*5*		*2*		3	41	0	23	12
R-55 (sc55)															
Field F1	*1*		21	30	18	25	*2*		*3*	*1*	3	56	44	45	57
Observed Parents		**A2**					***A1***					1	1	1	0
in vitro F1			23	11	0	0	*5*		*2*			44	0	27	11

*Outlined cells indicate the expected genotypes for the Field F1, in vitro F1, and parental isolates. Italic font and no outline indicate where observed data did not cohere with expectations.
**Where the haplotype designation was unclear, i.e. due to recombination, the isolate was classified as Unknown. These isolates are indicated in S9-11 Figs for ROIs 1, 4, and 5.
***Haplotype counts do not necessarily sum to the population size due to the Unknown isolates.

Table S4. Genotype and haplotype counts for regions of differentiation between the field F_1 and *in vitro* F_1.

Region	Scaffold	Minimum Significant Position (bp)	Maximum Significant Position (bp)	ROI size (bp)	Significant SNPs (#)	Significant SNPs in region (%)	Density of Significant SNPs (SNP/kb)	Linkage group
-	7	422,903	1,260,693	837,790	9	7.09	0.01	1 & 13
-	18	720,911	895,992	175,081	3	18.75	0.02	2
-	37	183,312	511,376	328,064	5	6.58	0.02	3
-	6	58,320	945,375	887,055	20	20.2	0.02	4
-	10	375,319	1,112,682	737,363	7	5.83	0.01	5
ROI-2	21	615,928	906,387	290,459	29	67.44	0.1	5
-	24	401,564	662,972	261,408	9	26.47	0.03	5
-	52	6,287	133,781	127,494	5	20	0.04	5
-	63	66,589	303,371	236,782	7	21.21	0.03	5
-	68	49,163	293,159	243,996	13	28.89	0.05	5
-	8	1,204	1,107,380	1,106,176	19	12.67	0.02	8
-	22	515,495	704,555	189,060	8	23.53	0.04	8
-	58	107,853	313,193	205,340	8	16.33	0.04	8
-	62	181,179	212,989	31,810	3	25	0.09	8
-	2	295,579	1,796,377	1,500,798	15	6.55	0.01	10 & 13
-	3	45,932	836,688	790,756	5	3.76	0.01	10 & 13
-	34	320,234	405,897	85,663	9	42.86	0.11	10
-	53	22,057	411,035	388,978	9	15.79	0.02	12
-	36	112,937	484,604	371,667	13	37.14	0.03	13
-	20	143,241	806,490	663,249	20	15.5	0.03	16
ROI-1	33	47,384	560,094	512,710	52	55.32	0.1	16
-	55	23,126	372,172	349,046	9	16.07	0.03	16

Table S5. Regions of differentiation between the field F_1 and inbred subpopulations.

	Genotype counts*							Haplotype counts			Haplotype (%)***		
	H1a/H1a	H3a/H4a	H1a/H3a	H1a/H4a	H3a/H3a	H4a/H4a	*Unknown***	**H1a**	**H3a**	**H4a**	**H1a**	**H3a**	**H4a**
Observed Parents	**A1**	**A2**						2.0	1.0	1.0	0.5	0.25	0.25
Field F1													
2009	2		15	15	*1*		3	34	17	15	47.22	23.61	20.83
2010			20	19				39	20	19	50.00	25.64	24.36
2011	1		15	12	*1*			29	17	12	50.00	29.31	20.69
Total	**3**		**50**	**46**	**2**		**3**	102	54	46	49.04	25.96	22.12
Field inbred													
2011	4		7	3				18	7	3	64.29	25.00	10.71
2012	6		6		4			18	14	0	56.25	43.75	0.00
2013	5	1	12	1	4			23	21	2	50.00	45.65	4.35
Total	**15**	**1**	**25**	**4**	**8**			59	42	5	55.66	39.62	4.72

*Outlined boxes denote the expected genotypes for the parents, Field F1 and Field inbred, assuming simple Mendelian inheritance.

**Where haplotype was unclear, i.e. due to recombination between distinct haplotypes, the isolate was classified as Unknown. These isolates are indicated in S10B and S14 Fig

***Haplotype frequencies do not necessarily sum to one due to the Unknown genotypes.

Table S6. Genotype and haplotype counts in ROI-1 in the field population with respect to year and subpopulation (F_1 vs. inbred).

	Genotype counts*							Haplotype counts			Haplotype (%)***		
	H1/H1	H3/H4	H1/H3	H1/H4	H3/H3	H4/H4	*Unknown***	**H1**	**H3**	**H4**	**H1**	**H3**	**H4**
Observed Parents	**A1**	**A2**						2	1	1	0.5	0.25	0.25
Field F1													
2009			21	15				36	21	15	50.00	29.17	20.83
2010			24	15				39	24	15	50.00	30.77	19.23
2011			21	8				29	21	8	50.00	36.21	13.79
Total			**66**	**38**				104	66	38	50.00	31.73	18.27
Field inbred													
2011	9		3	1			1	22	3	1	78.57	10.71	3.57
2012	9		4	3				25	4	3	78.13	12.50	9.38
2013	10	5	1	6			1	27	6	11	58.70	13.04	23.91
Total	**28**	**5**	**8**	**10**			2	74	13	15	69.81	12.26	14.15

*Outlined boxes denote the expected genotypes for the parents, Field F1 and Field inbred, assuming simple Mendelian inheritance.

**Where haplotype was unclear, i.e. due to recombination between distinct haplotypes, the isolate was classified as Unknown. These isolates are indicated in S10B and S14 Fig

***Haplotype frequencies do not necessarily sum to one due to the Unknown genotypes.

Table S7. Genotype and haplotype counts in ROI-2 in the field population with respect to year and subpopulation (F_1 vs. inbred).

Scaffold	Minimum Position (bp)	Maximum Position (bp)	Size (bp)	Number of SNPs in region	Total SNPs in scaffold
Field F1					
2	857,230	1,027,986	170,756	17	276
4	78,389	936,543	858,154	69	258
27	1,260	559,598	558,338	80	156
34	373,020	419,673	46,653	14	90
40	53,533	62,968	9,435	4	92
Field inbred					
4	204,373	896,170	691,797	53	258
27	8,102	506,357	498,255	20	156

Table S8. Regions of differentiation between isolates of opposite mating types in the field F_1 and inbred subpopulations.

APPENDIX C

SUPPLEMENTARY FIGURES

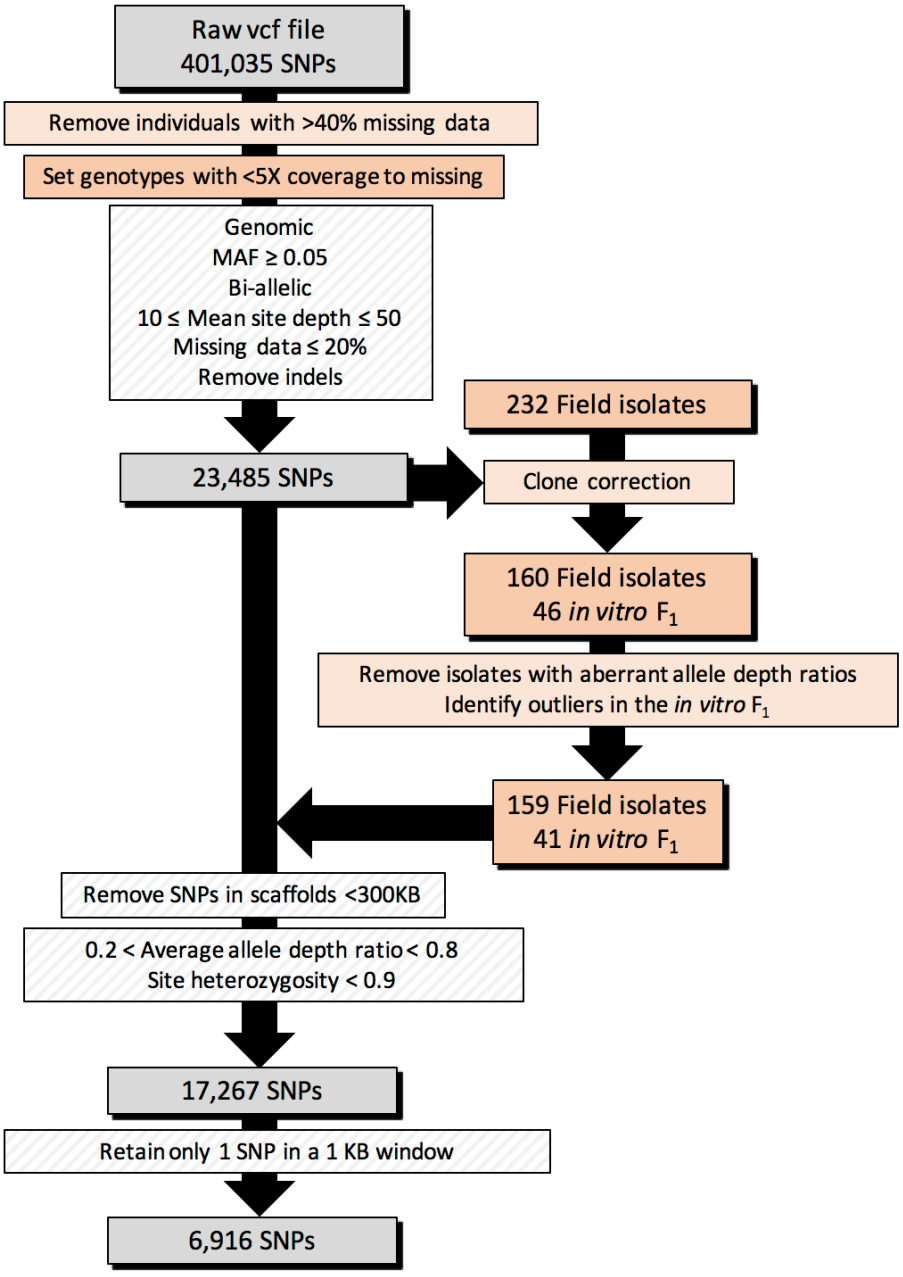


Figure S1. SNP and individual filtering pipeline. Gray and white shaded boxes indicate SNP filtering steps. Orange shaded boxes indicate individual filtering steps.

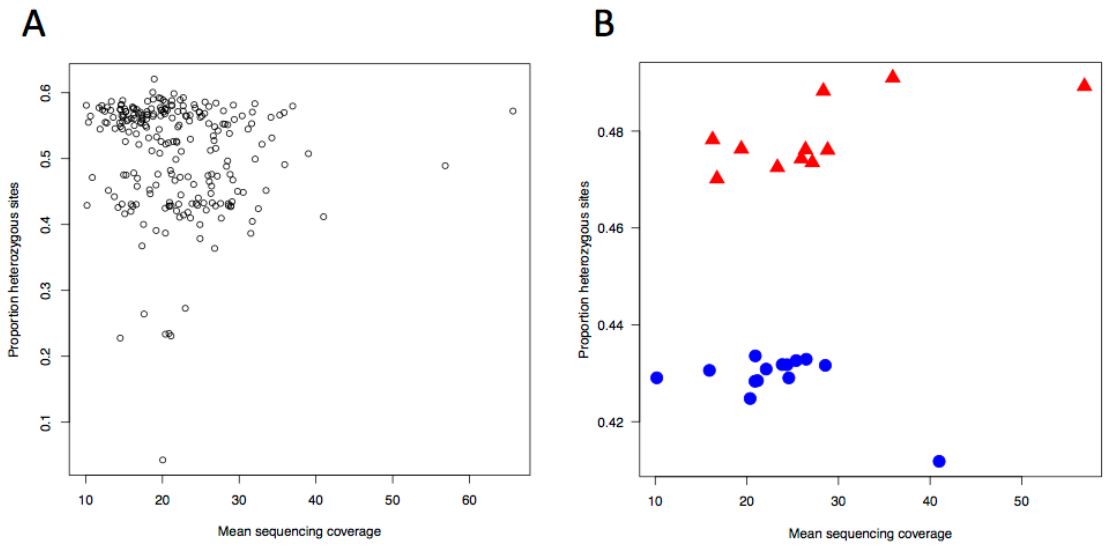


Figure S2. Relationship between sequencing coverage and heterozygosity.
The proportion of heterozygous genotypes per sample plotted against individual mean sequencing coverage (n=23,485 SNPs). A) For all samples prior to clone-correction and outlier removal, and B) for only replicates of the A1 parental isolate (n=14; blue circles) and the A2 parental isolate (n=11; red triangles).

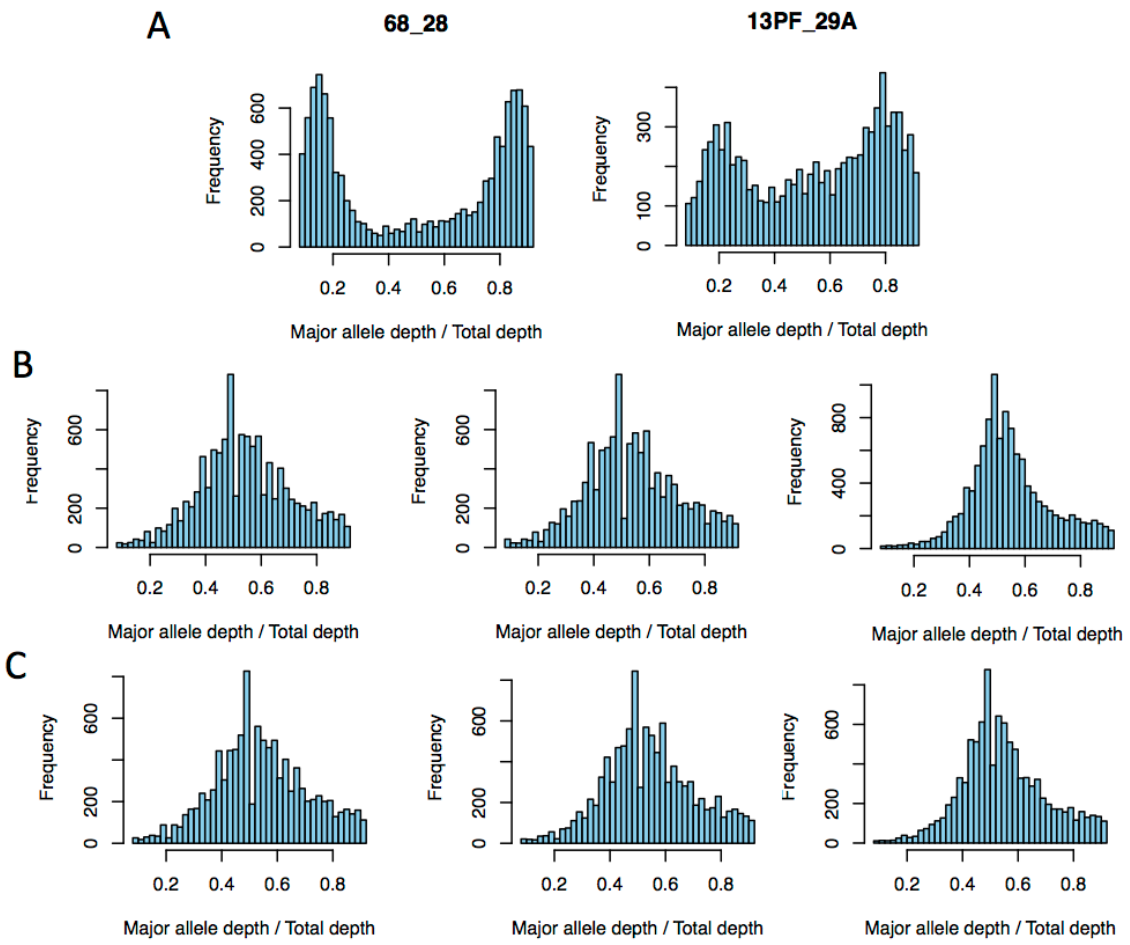


Figure S3. Skewed allele depth ratios in two isolates suggest ploidy variation. Histograms of the ratio of the major allele depth to the total depth for each heterozygous genotype for each isolate (at 23,485 SNPs). A) One *in vitro* and one field isolate display grossly aberrant allele depth ratios suggestive of ploidy variation. In contrast, allele depth ratios for the (B) A1 and (C) A2 parental isolates were centered at approximately 0.5.

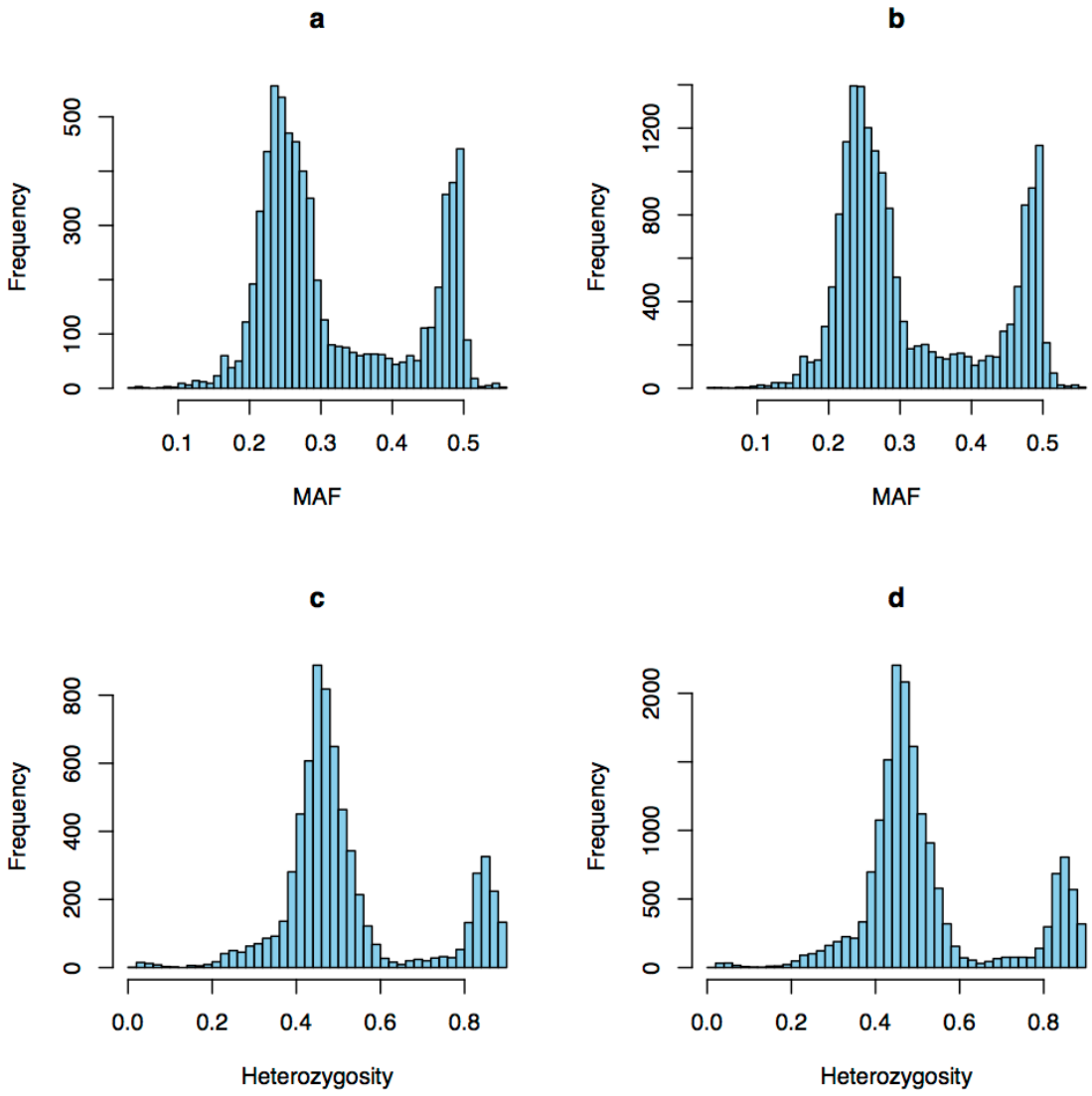


Figure S4. Comparing pruned and unpruned data sets. Minor allele frequency (MAF) and heterozygosity distributions for the unpruned (n=17,267) and pruned SNPs (n=6,916) in the field population (n=159 isolates).
(A) and (C) are for the pruned data set. (B) and (D) are for the unpruned data set.

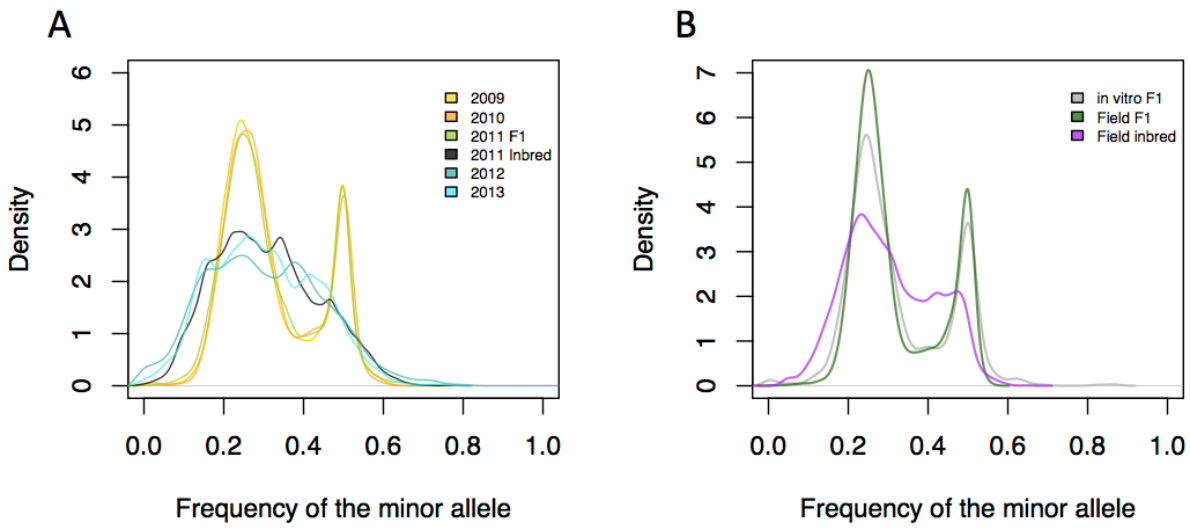


Figure S5. Minor allele frequency (MAF) distributions for the field and *in vitro* populations. A) MAF distributions for each year in the field population. Year 2011 was split into F_1 and inbred isolates based on classification via Mendelian errors, showing that the 2011 F_1 contingent MAF distribution was similar to that of 2009 and 2010, which contained exclusively F_1 isolates. Within years containing F_1 isolates, we observe peaks at 0.25 and 0.5, consistent with expectations for a population derived from only two parents. B) The field F_1 subpopulation MAF distribution was consistent with the that of the *in vitro* F_1. The field Inbred MAF distribution deviated from expectations for an F_1 suggesting incidence of allele frequency changes.

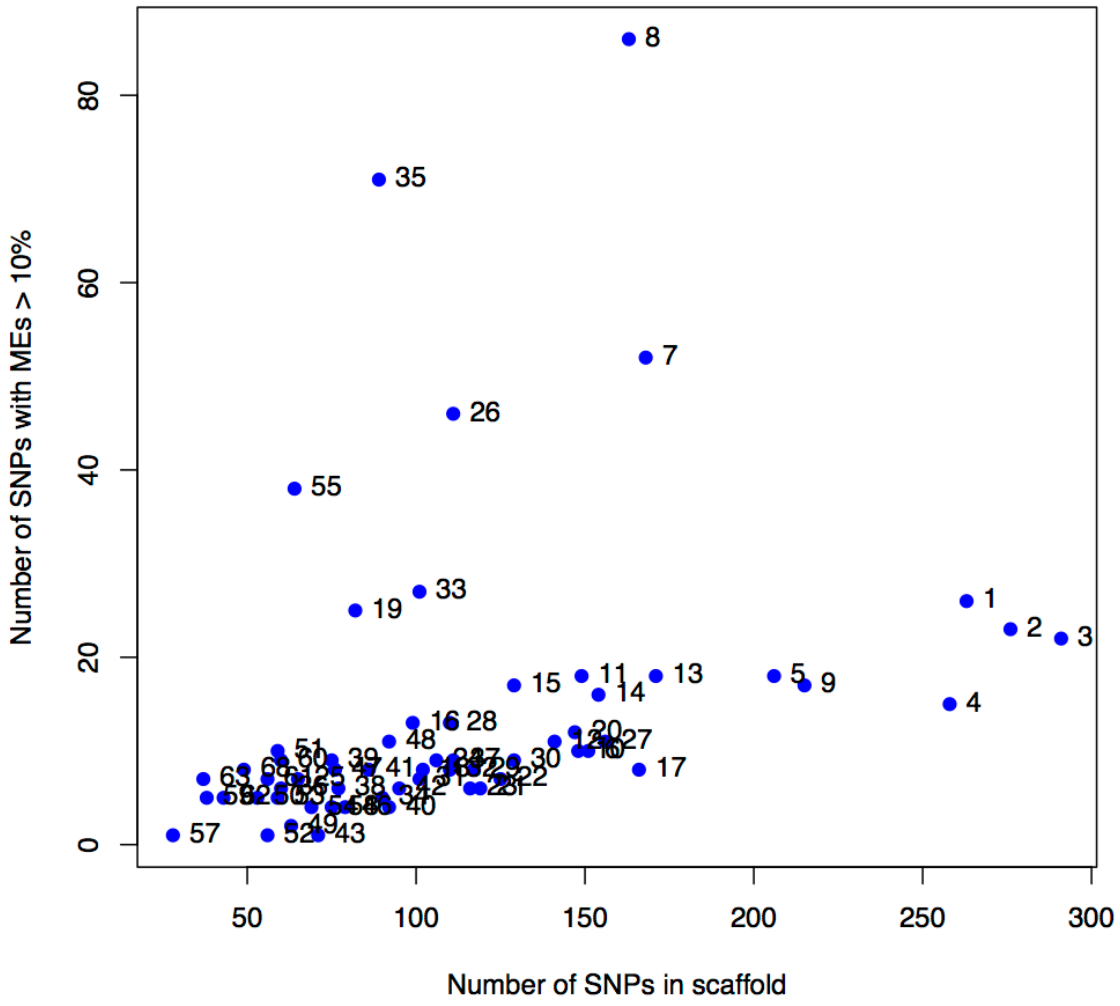


Figure S6. Relationship between number of SNPs in each scaffold and the incidence of Mendelian error (ME) enriched SNPs among the *in vitro* F_1 and empirically defined field F_1 (n=143) prior to removal of the ME enriched SNPs. SNPs enriched for MEs were defined as SNPs where greater than 10% of *in vitro* F_1 and field F_1 isolates had a ME (at least 15 isolates). The number of ME enriched SNPs was plotted as a function of the number of SNPs in each scaffold, identifying seven scaffolds (7, 8, 19, 26, 33, 35, and 55) with excess ME-enriched SNPs relative to the other scaffolds. Data points are labeled with the scaffold number.

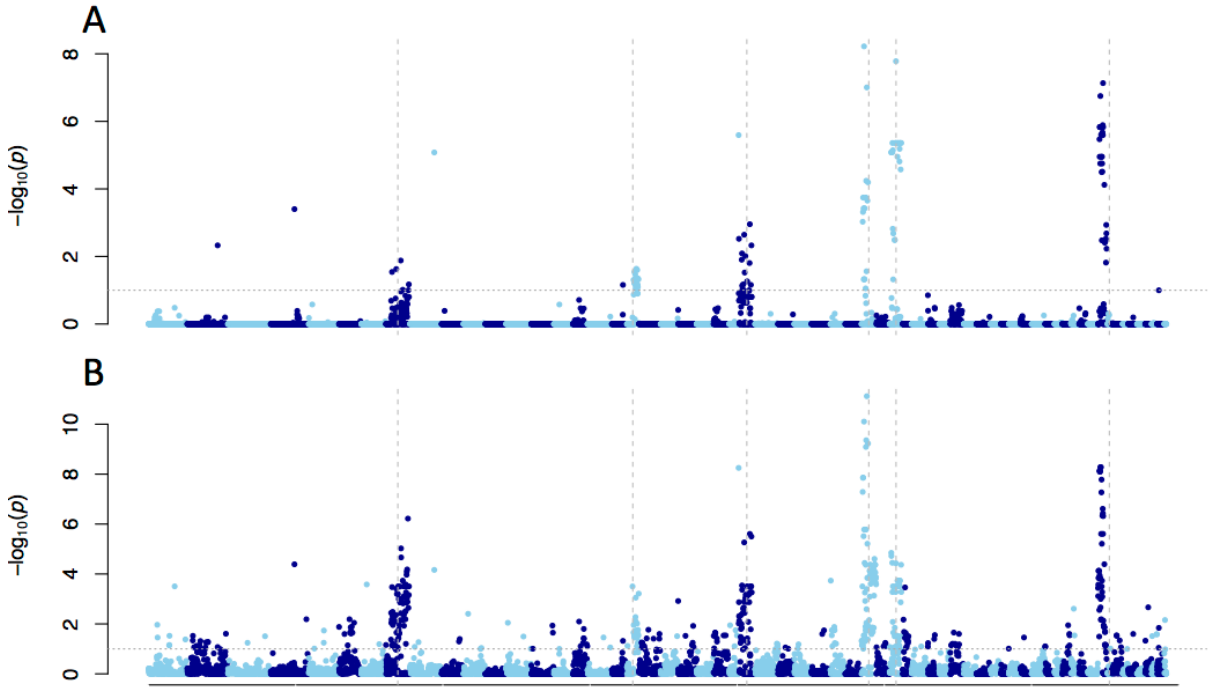


Figure S7. Regions of differentiation between the *in vitro* F_1 and the field F_1 and field inbred subpopulations identified using Fisher's exact tests of allele frequency differences at each SNP. Negative log10-transformed, false-discovery rate (FDR) adjusted, *P*-values from pairwise comparisons between the (A) *in vitro* F_1 and field F_1 and (B) *in vitro* F_1 and field inbred plotted for each SNP. SNPs are ordered relative to physical position and colors alternate by scaffold. Gray vertical dashed lines in A-C indicate scaffolds pertaining to regions of differentiation between the *in vitro* F_1 and the field F_1. The gray dotted line in A and B denotes the 10% FDR threshold.

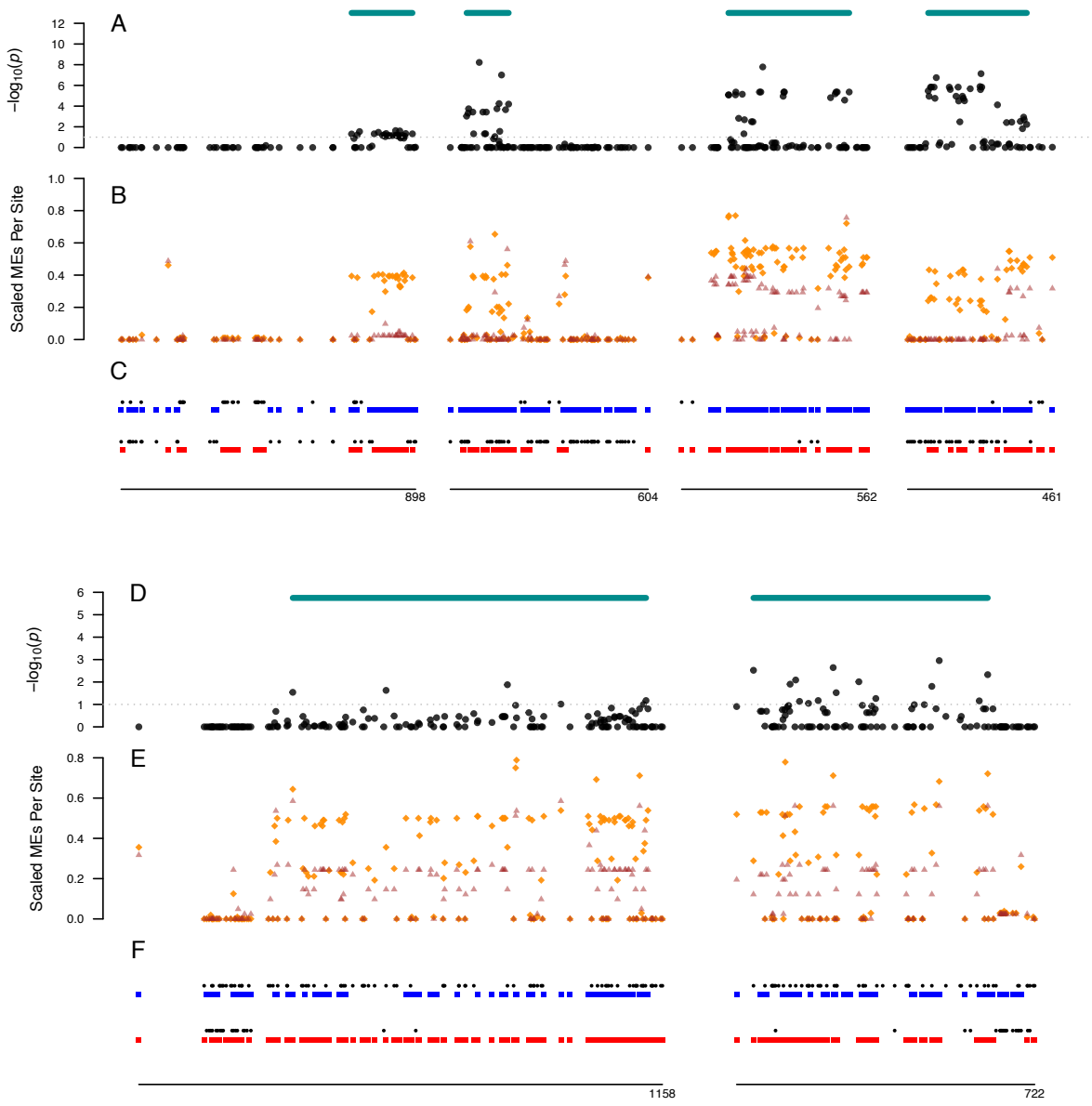


Figure S8. Regions of differentiation between the *in vitro* F_1 and field F_1 were associated with loss of heterozygosity (LOH) events in the parental cultures. A) and D) show the negative $\log_{10}$-transformed, FDR adjusted, *P*-values from the Fisher's exact test of allele frequency differences between the *in vitro* F_1 and field F_1, relative to physical position (kb), in scaffolds corresponding regions of differentiation. The teal bars span each differentiated region. B) and E) show the proportion of individuals with a Mendelian error (ME) for each SNP in the *in vitro* F_1 (brown triangles) and the field F_1 (orange diamonds), excluding homozygous isolates. C) and F) are the parental genotypes represented by blue (A1 parent) and red (A2 parent) squares for homozygous genotypes and black dots for heterozygous genotypes.

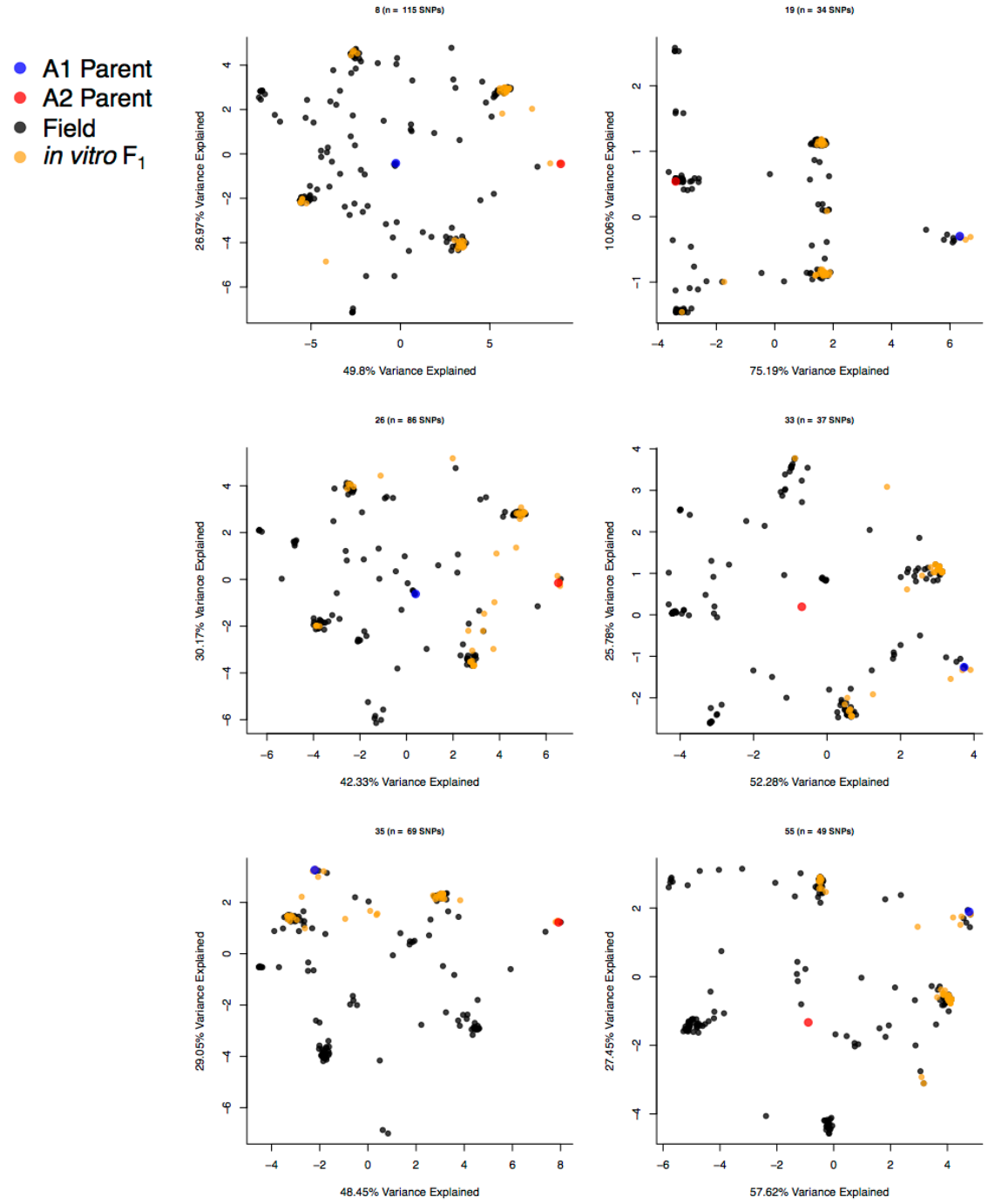


Figure S9. Principal component analysis (PCA) in scaffolds pertaining to regions of interest (ROIs). PCA was performed on the in vitro F_1, field F_1, and the field inbred isolates, as well as, the consensus parental genotypes with only SNPs in each of the six differentiated regions. All PCAs show four primary clusters, corresponding four genotypic classes. The field isolates (n=159) are represented by closed, black circles. The *in vitro* F_1 (n=41) are represented by orange, closed circles. The A1 and A2 consensus parental genotypes are represented by blue and red closed circles, respectively.

A

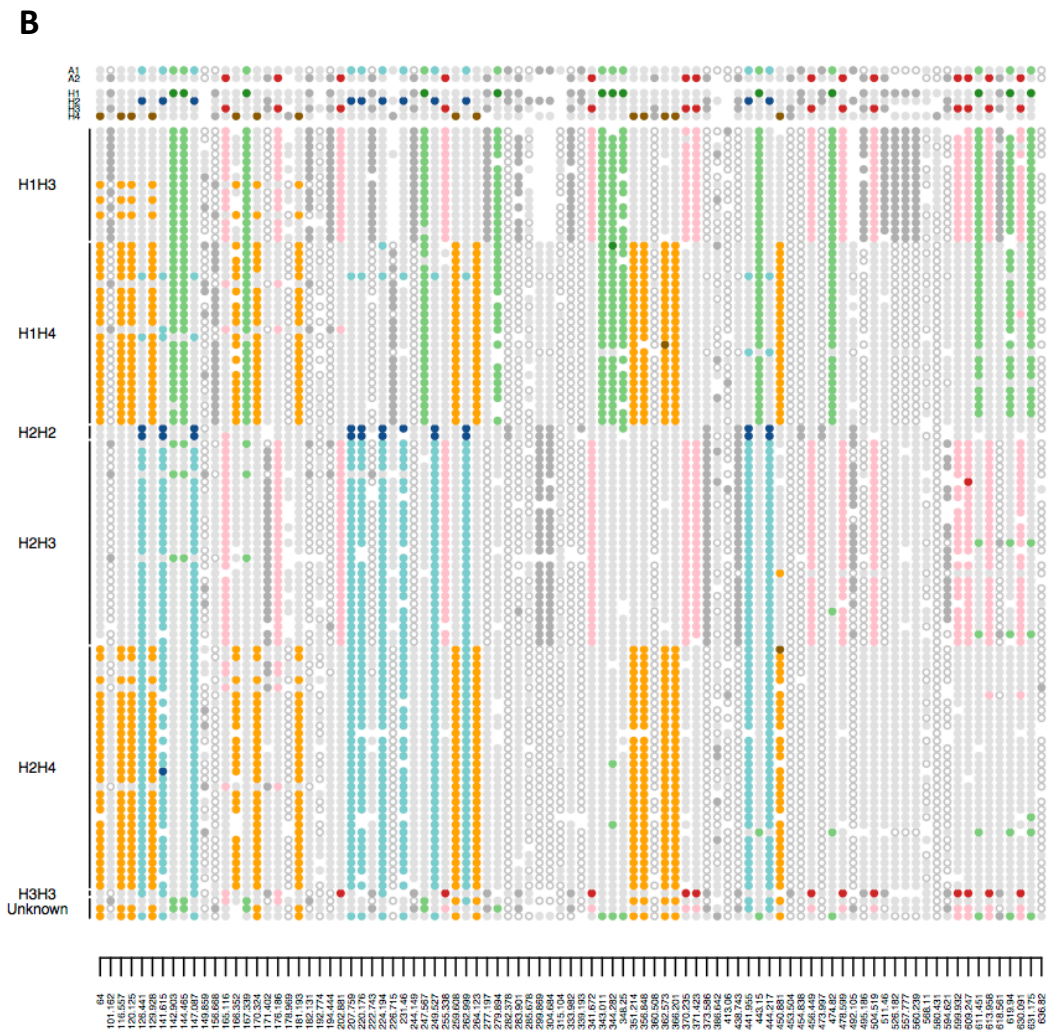
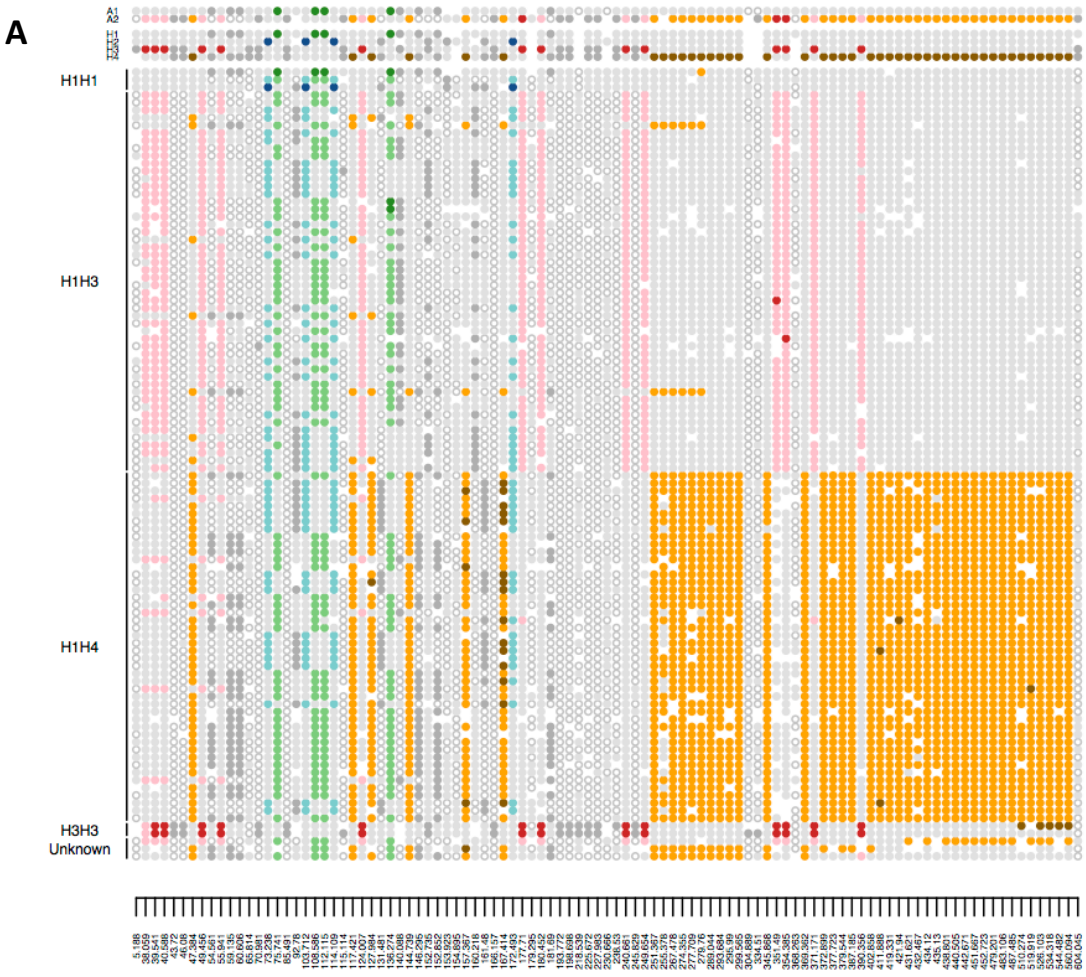


Figure S10. Phase diagram for R-26 (scaffold 26)
Haplotype tagging SNPs were identified from the phased parental genotypes, and each isolate genotype was represented with respect to these tagging SNPs (see Methods). Closed, colored circles indicate haplotype tagging SNPs, with darker colors indicating the homozygote condition, and lighter shades indicating the heterozygous condition. Filled, gray circles indicate homozygous genotypes at non-tagging SNPs, whereas open, gray circles indicate heterozygous genotypes at non-taggging SNPs (see legend). Missing genotypes are denoted by the absence of a circle. Phase diagrams for the parental isolates (top), identified haplotypes (middle) and (A) *in vitro* F_1 and (B) field F_1.


A
A1
A2
H1H1
H1H3
H1H4
H3H3
Unknown

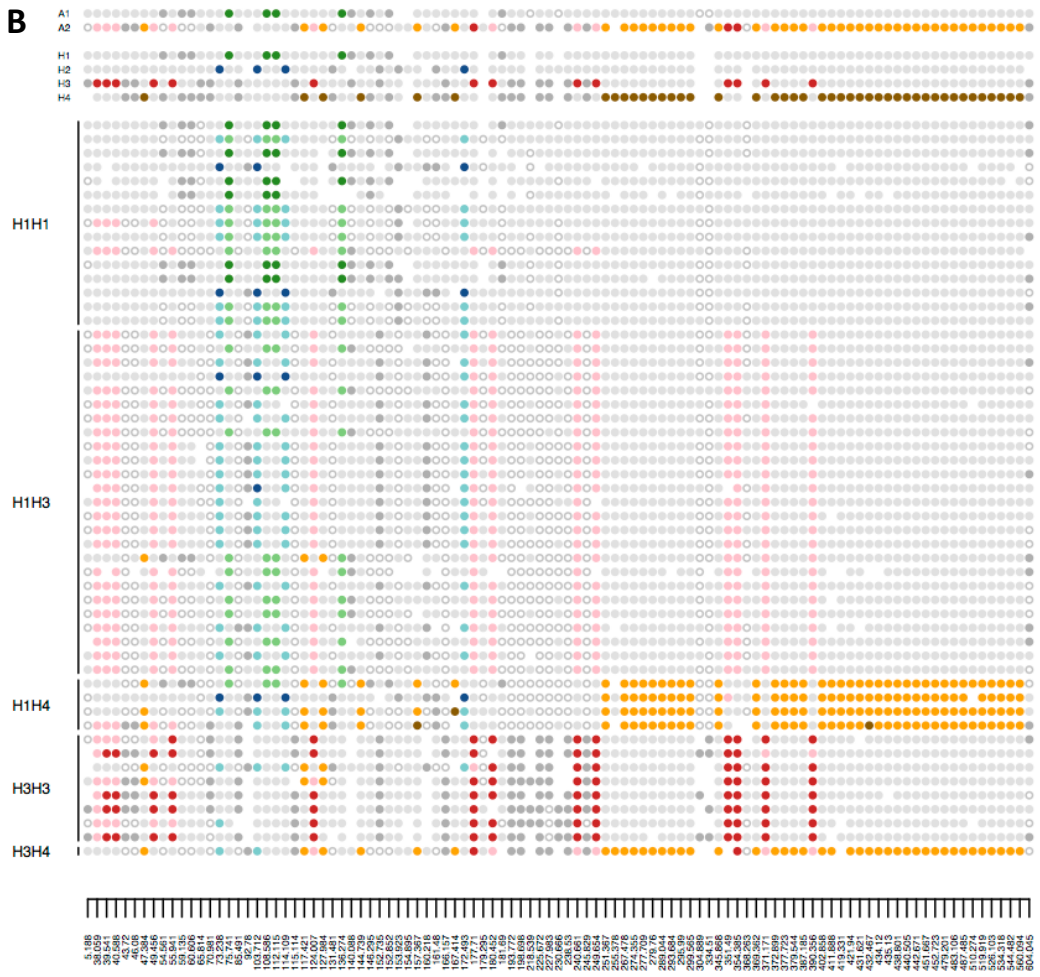
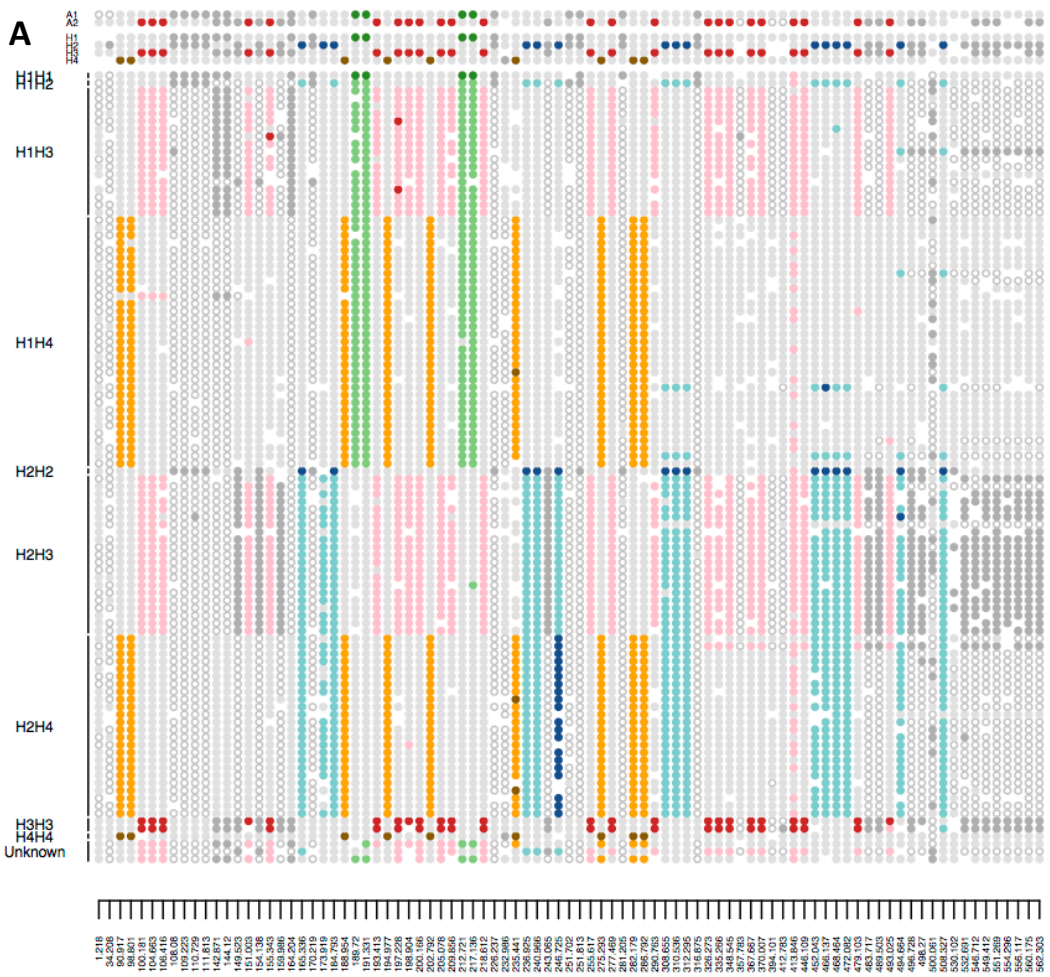


Figure S11. Phase diagram for ROI-1 (scaffold 33)
Labeled genotypes based on ROI-1 only. Haplotype tagging SNPs were identified from the phased parental genotypes, and each isolate genotype was characterized with respect to the tagging SNPs (see Methods). Closed, colored circles indicate haplotype tagging SNPs, with darker colors indicating the homozygote condition, and lighter shades indicating the heterozygous condition. Filled, gray circles indicate homozygous, non-tagging genotypes, whereas open, gray circles indicate heterozygous non-tagging genotypes (see legend). Missing genotypes are denoted by the absence of a circle. Phase diagrams for the parental isolates (top), identified haplotypes (middle) and (A) field F_1 and (B) field inbred (bottom).


A
A1
A2
H1H1
H1H2
H1H3
H1H4
H2H2
H2H3
H2H4
H3H3
H4H4
Unknown

B

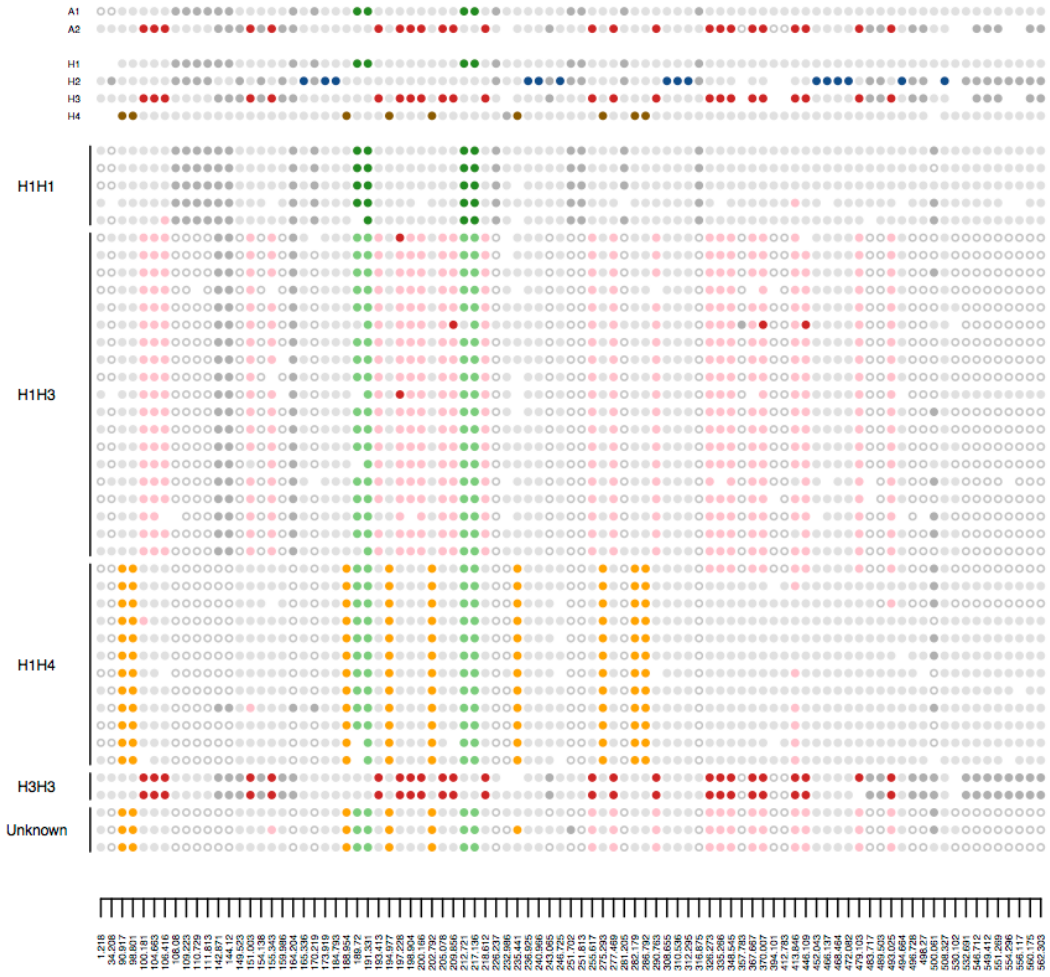


Figure S12. Phase diagram for ROI-5 (scaffold 35). Haplotype tagging SNPs were identified from the phased parental genotypes, and each isolate genotype was characterized with respect to the tagging SNPs (see Methods). Closed, colored circles indicate haplotype tagging SNPs, with darker colors indicating the homozygote condition, and lighter shades indicating the heterozygous condition. Filled, gray circles indicate homozygous, non-tagging genotypes, whereas open, gray circles indicate heterozygous non-tagging genotypes (see legend). Missing genotypes are denoted by the absence of a circle. Phase diagrams for the parental isolates (top), identified haplotypes (middle) and (A) *in vitro* F_1 and (B) field F_1.

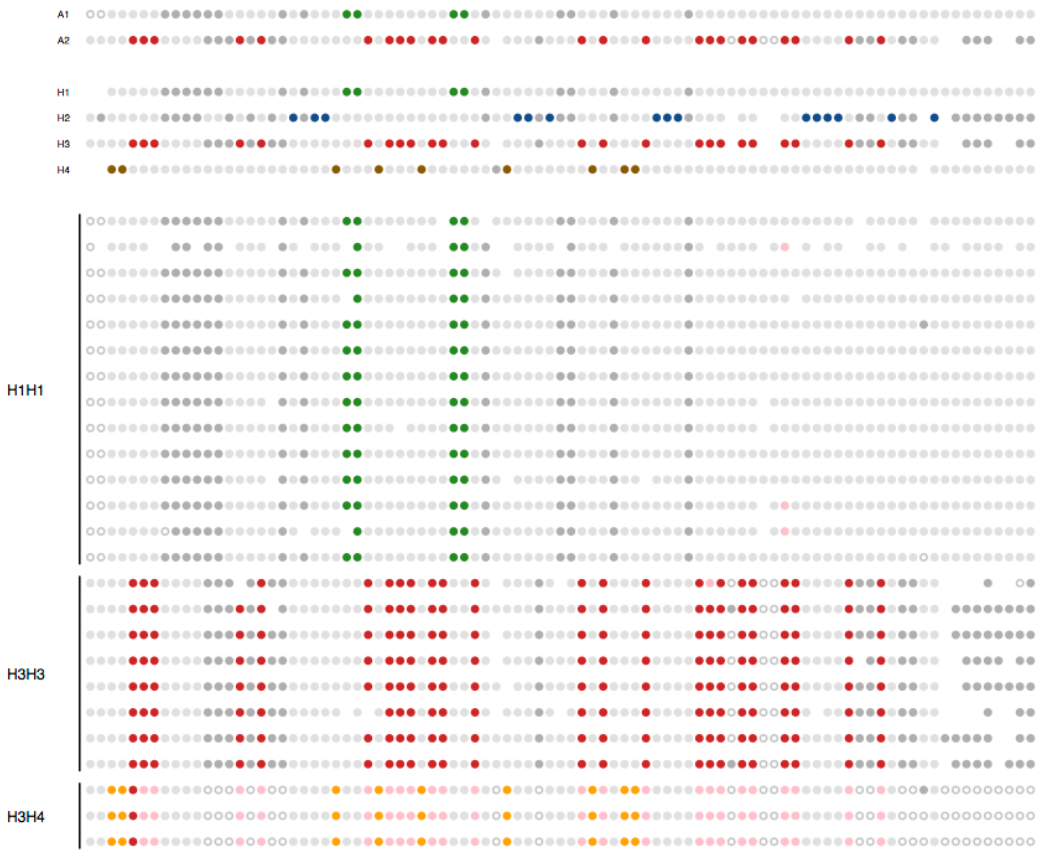


Figure S13. Phase diagrams for the parental replicates. All A1 parental replicates were H1/H1. The three A2 parental replicates sequenced prior to XX were H3/H4, whereas the later sequenced replicates were H3/H3 (Table B.S1).

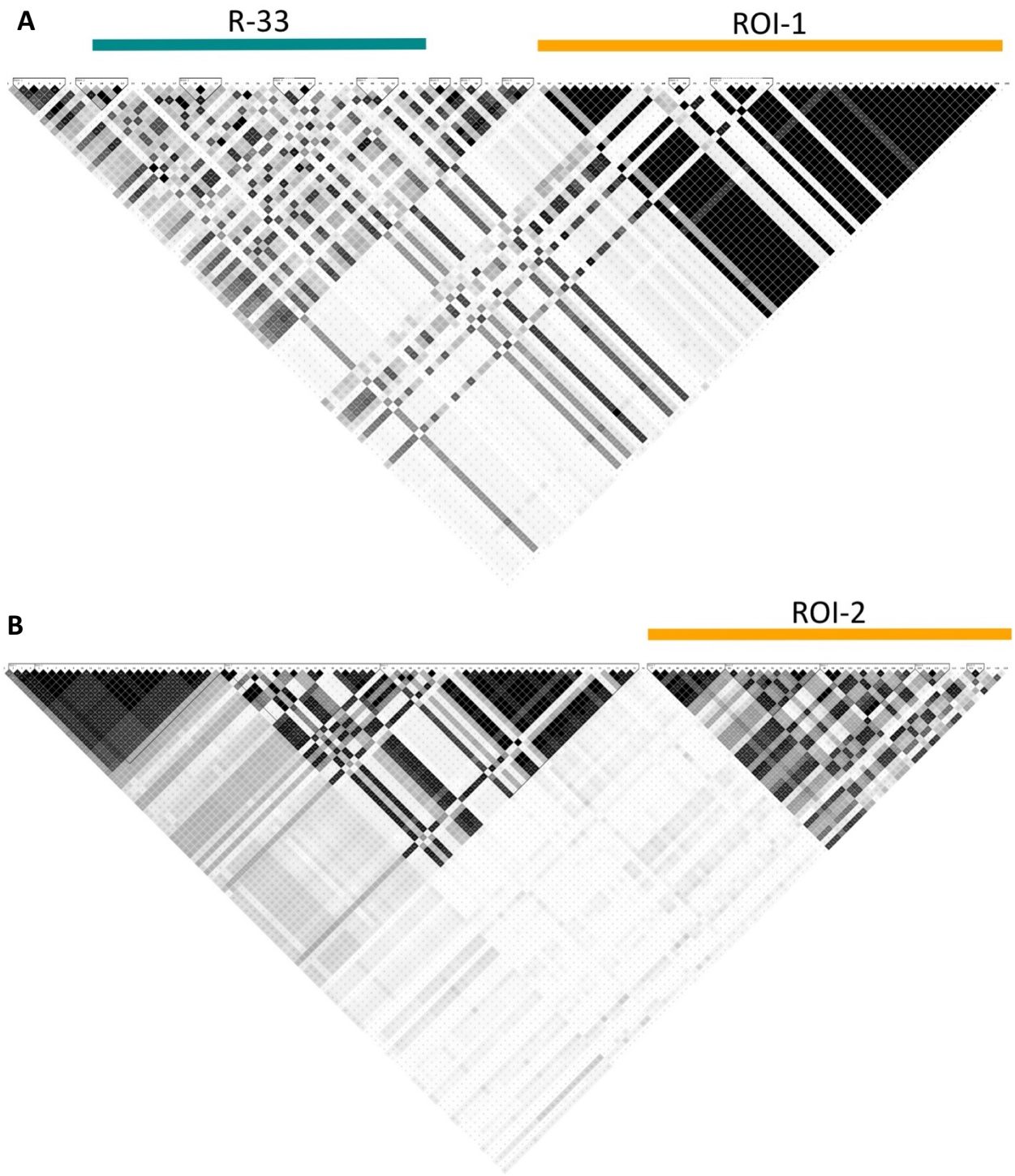


Figure S14. Linkage disequilibrium (LD) in regions of interest (ROIs), where ROIs are indicated by orange bars. A) LD in scaffold 33 which contains both R-33 and ROI-1. B) LD in scaffold 26, which contains ROI-2.

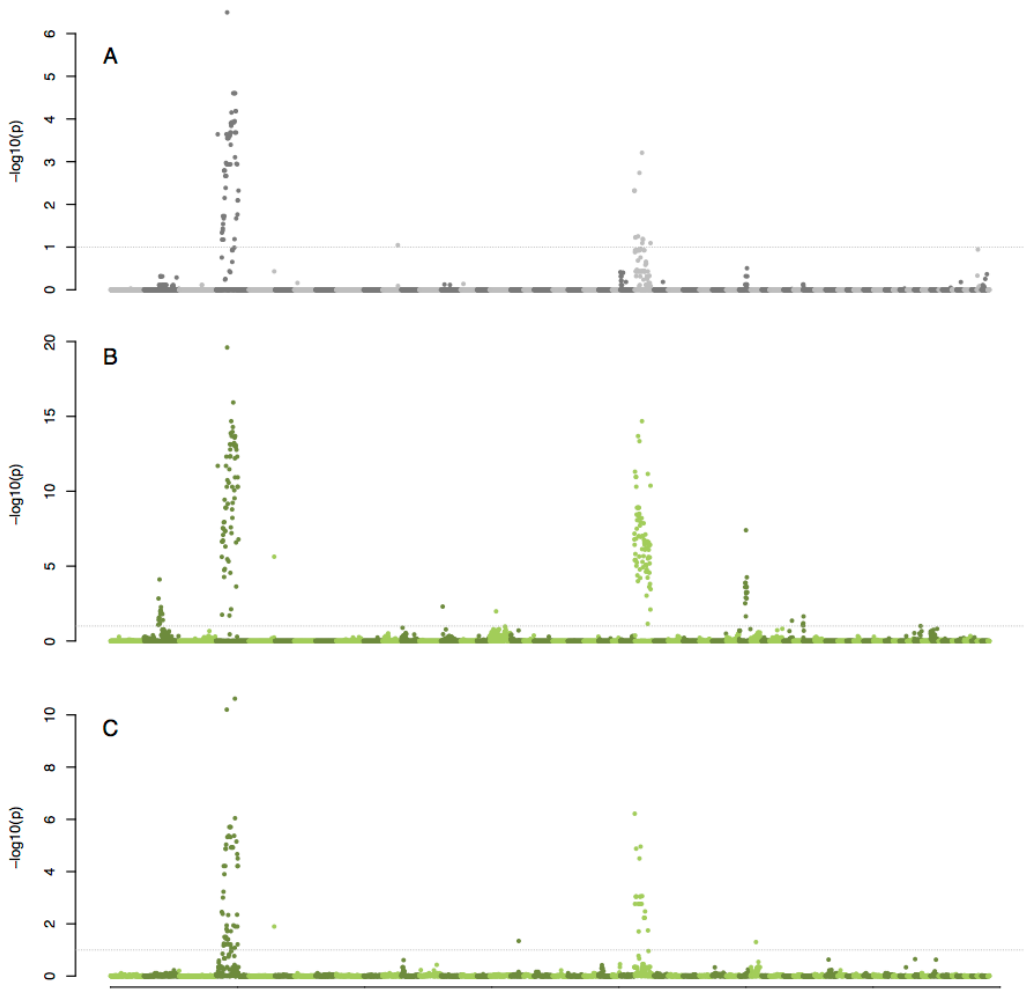


Figure S15. Allele frequency differences between isolates of opposite mating types in the *in vitro* F_1, field F_1, and field inbred subpopulations. Negative $\log_{10}$-transformed, FDR corrected *P*-values ordered by scaffold and physical position, from the Fisher's exact test of allele frequency differences between A1 and A2 isolates in the (A) *in vitro* F_1, (B) field F_1, and (C) field inbred subpopulations. SNPs above the gray lines in A-C were significant at a 10% false-discovery rate (FDR) threshold.

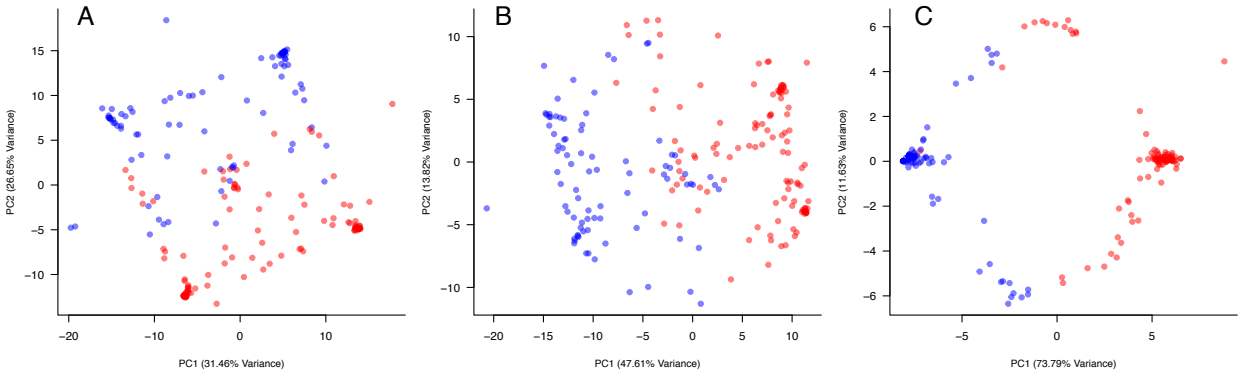


Figure S16. Principal component analysis in the mating type region (MTR). PCA of all *in vitro* and field isolates using the: A) 293 SNPs in the MTR; B) 184 significantly differentiated SNPs in the field F_1; and the C) 51 SNPs significantly differentiated in both the field F_1 and inbred subpopulations.

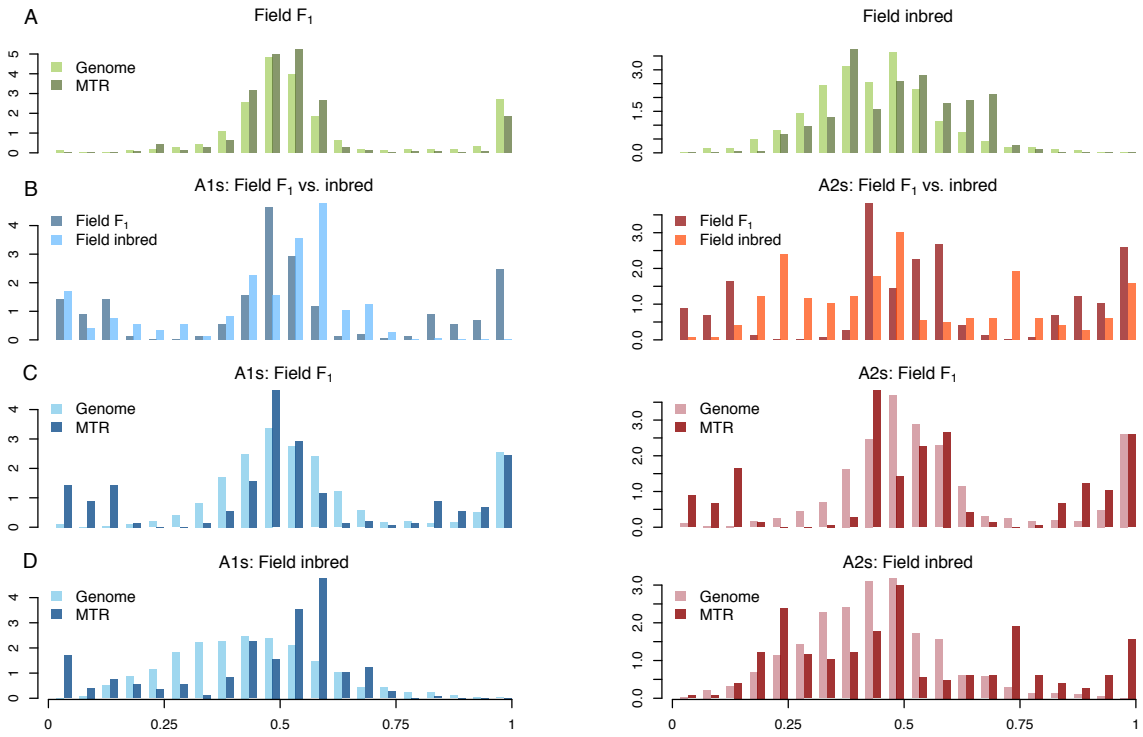


Figure S17. Heterozygosity in the mating type region (MTR) compared to the rest of the genome. The distribution of heterozygosity in the MTR relative to the genome for the: A) field F_1 and field inbred isolates; B) A1 and A2 field F_1 vs. field inbred isolates in the MTR; C) A1 and A2 field F_1 isolates in the MTR relative to the genome; and D) A1 and A2 field inbred isolates in the MTR relative to the genome.

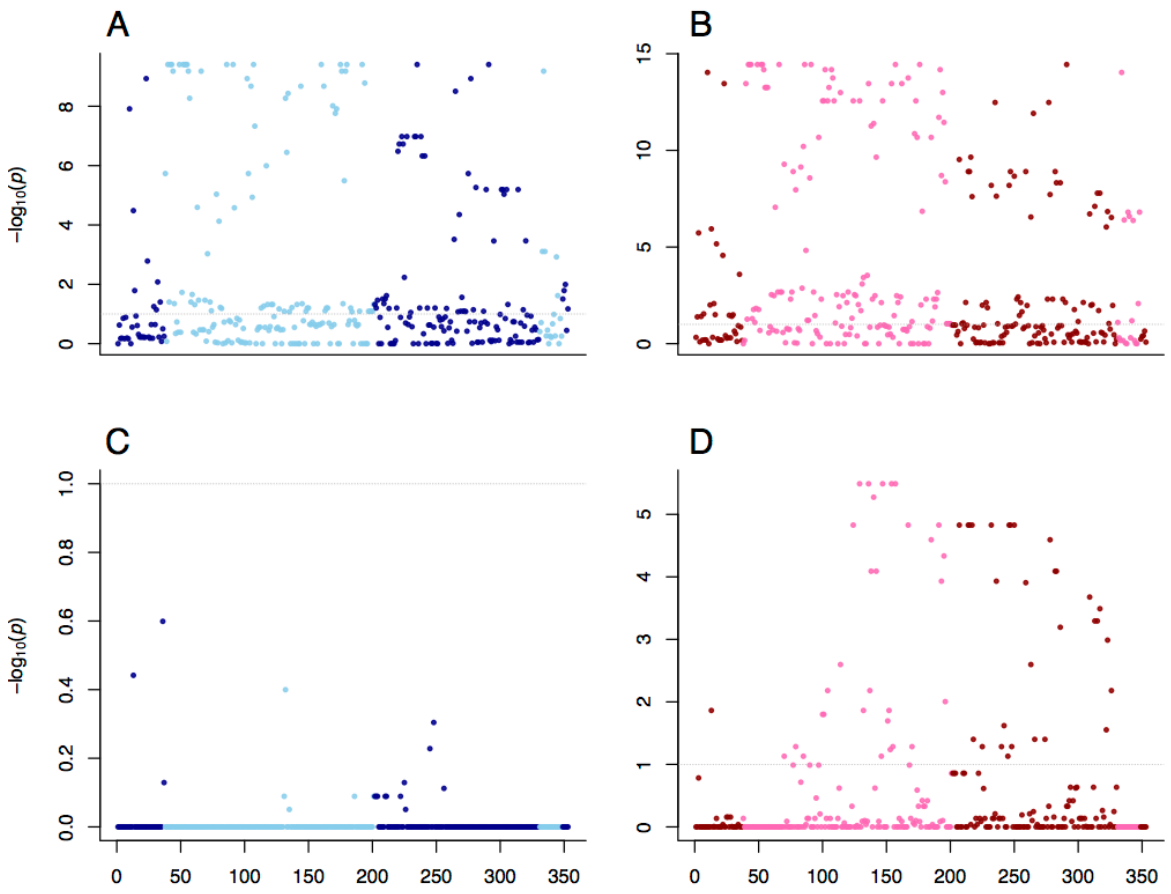


Figure S18. Exact test of heterozygote excess in all five mating type associated sub-regions (*n*=353 SNPs), ordered by position in scaffolds 2, 4, 27, 34, and 40. A) A1 field F_1 isolates. B) A2 field F_1 isolates. C) A1 field inbred isolates. D) A2 field inbred isolates.